Charles University

Faculty of Arts

Doctoral Dissertation

Mgr. Jakub Jehlička

Gesture and Eventuality
A Crosslinguistic Study

Department of Linguistics

Supervisor:	doc. Mgr. Josef Fulka, Ph.D.
Advisor:	Prof Dagmar S. Divjak
Study programme:	General Linguistics

Prague 2021

Prohlašuji, že jsem disertační práci napsal samostatně s využitím pouze uvedených a řádně citovaných pramenů a literatury a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne ..... podpis autora

Abstract:
Speakers of typologically distinct languages were shown to comprehend, or even perceive, the same events differently, constrained by the specific grammatical means for encoding event structure available in their languages. Across languages, co-speech gestures (i.e. bodily movements that, in an orchestrated manner, co-occur with speech) were also observed to be used in different ways that reflect the language-specific embodied patterns of event conceptualization. Together, these kinds of the "weak linguistic relativity" effects are sometimes referred to as "thinking, perceiving and gesturing for speaking".

This thesis, following on from previous studies of the gesture-grammar interface across languages, deals with multimodal construals of events in Czech and English. Specifically, it explores the association of gestural formal features (movement manner and ending) and semantic features that constitute an aspectual contour of an event (i.e. the construal of the temporal and qualitative unfolding of an event).

First, a corpus-based study was carried out, analysing material obtained from Czech and English multimodal corpora. The material consisted of recordings of spontaneous language production captured in an interactional setting (academic business meetings). The quantitative analysis (random forests of conditional inference trees) revealed that in English, gestures with a marked ending are mostly associated with achievement predicates and complex gestural forms often co-occur with progressive verb forms. In Czech, ended gestures had a stronger association with the finer-grained semantic feature of directedness, and complex forms were observed to accompany events with incremental construals (these features were, however, also strongly associated with perfective and imperfective aspect, respectively).

Subsequently, a behavioural experiment was carried with Czech speakers, focused on the comprehension of multimodal patterns observed in the corpus study. The aim of the experiment was to assess to what degree speakers associate the presence or absence of the gestures' marked ending feature with aspect. The results indicated a strong association between perfective predicates and ended gestures.

In both languages, gestures take part in the construals of events – in some cases, gestures even provide critical information needed for discrimination between alternate construals. The observed differences in Czech and English multimodal eventuality expressions can be attributed to the differences in the grammatical encoding of aspectuality that constrain the ways in which gestures are used for semantic profiling.

Keywords:
aspectuality, cognitive linguistics, construction grammar, construal, co-speech gesture, Czech, English, event cognition, eventuality, iconicity, language interaction, multimodal corpora, multimodality

Abstrakt:
Mluvčí typologicky odlišných jazyků volí různé strategie při popisu stejné události v závislosti na dostupných jazykově-specifických gramatických prostředcích. Tyto strategie se projevují např. různými způsoby konceptualizace událostních rámců během jazykového vyjádření, ale v nejazykové kognici. Jedním z jevů, které byly v této souvislosti zaznemány, jsou jazykově specifické způsoby gestikulace doprovázející mluvené popisy událostí, které reflektují (či manifestují) tělesně ukotvená konceptuální schémata, na nichž naše vnímání událostí stojí.

Tématem této práce je multimodální konstruování (*construal*) událostí v češtině a v angličtině. Konkrétně se práce zaměřuje na spojitosti mezi formálními rysy gest (způsob pohybu a jeho zakončení) a sémantické rysy, které konstituují tzv. aspektuální kontury událostí (konstruování časového a kvalitativního průběhu události).

První část prezentovaného výzkumu tvoří analýza materiálu z českého a anglického multimodální korpusu. Oba použité korpusy obsahují nahrávky spontánních projevů v interakcích zachycených během pracovních jednání v akademickém prostředí. Kvantitativní analýzy (metoda tzv. klasifikačních stromů a náhodných lesů) ukázala, že a) v angličtině je významným prediktorem výskytu gest s rysem ukončenosti aktionsartová kategorie *achievement* (telické okamžité děje) a že progresivní slovesné tvary predikují výskyt komplexních gestických forem, a b) že v češtině souvisí ukončenost gest s rozdílem mezi směrovými a nesměrovými typy událostí, přičemž nejvýznamnějším prediktorem výskytu komplexních gestických forem je sémantický rys inkrementálnosti. V češtině se zároveň ukázala silná korelace mezi aktionsartovými podtypy a gramatickým videm.

Druhou částí výzkumu je experimentální studie s českými mluvčími zaměřený na percepci některých typů multimodálních vzorů pozorovaných v českém korpusu. Účastníkům experimentu byla prezentována ukončená a neukončená gesta spolu s vidovými variantami vět, přičemž testované osoby volily, která z vět patří k danému gestu. Experiment ukázal silnou tendenci českých mluvčí spojovat perfektivní vid s ukončenými gesty.

V češtině i v angličtině je gestikulace důležitou součástí konstruování událostí – v některých případech dokonce gesta přinášejí kontext nutný k úspěšnému rozlišení alternativních způsobů konstruování. Rozdíly v multimodálních konstrukcích událostních rámců v češtině a angličtině odrážejí rozdíly v gramatickém značení aspektuálnosti, které podmiňují užívání gest za účelem sémantického profilování událostních rámců.

Klíčová slova:
angličtina, aspektuálnost, čeština, gestikulace, ikoničnost, kognitivní lingvistika, konstrukční gramatika, konstruování, multimodální korpusy, multimodálnost, sémantika událostí

Acknowledgements

I am endlessly grateful for all the help and support I received from a considerable number of people (and institutions as well).

First and foremost, let me thank **Josef Fulka**, my supervisor, for his patience, inspiration and all our sessions over a pint at The Rudolfinum. Where would we, the Prague gesture lot, be without him?

For her kind support and for setting me on the right track, I am indebted to **Dagmar Divjak**, my advisor and host supervisor during my year in Sheffield.

Special thanks go to **Eva Lehečková**, a fellow seeker of lost time. If this thesis is worth anything, it is thanks to my collaboration with her.

My thanks also go to (in alphabetical order):

> Shanley Allen, Jakub Čapek, Jan Chromý and Ercel group, Alan Cienki, Tomáš Doischer, Ondřej Dufek, Viktor Elšík, Mirjam Fried and Epocc group, James Hill, Jiří Januška, Ivan Kafka, Simon Kaiser, Magdalena Kersting, Jonathan Kilgour, Emina Kurtić, Michal Láznička, Claire Leavitt, Markéta Lisá, Katharina Lutz, Štěpán Matějka, Srdan Medimorec, Petar Milin, Irene Mittelberg, Hana Prokšová, Gudrun Rohde, Martin Sedláček, Jakub Sláma, Mark Turner, Gareth Walker and Jordan Zlatev.

I hope those I have forgotten will forgive me.

This thesis would not have been possible if it had not been for the help of the Institute of Deaf Studies, Charles University, that provided me with the necessary recording equipment and facilities.

Last, but not least, let me express my gratitude to the funding bodies, the Anglo-Czech Educational Fund and the European Commission, for their generous grants that made it possible for me to work on the dissertation project in the UK (Sheffield) and in Germany (Aachen). This work was also supported by the European Regional Development Fund project "Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World" (reg. no.: CZ.02.1.01/0.0/0.0/16_019/0000734), by the project "International Mobility of Researchers at Charles University" (CZ.02.2.69/0.0/0.0/16_027/0008495) and by the project "Progres Q10, Language in the shiftings of time, space, and culture".

I am solely responsible for any errors in this work.

> In Prague, Sheffield, Aachen and St. Georgenthal, 2014–2020.

Contents

1. Introduction

> *"It is often hard for the literate world to remember that the core ecology for language use is in face-to-face interaction – this is the niche in which languages are learnt and where the great bulk of language use occurs. In this niche, language production always occurs with the involvement of not only the vocal tract and lungs, but also the trunk, the head, the face, the eyes and, normally, the hands"* (Levinson and Holler, 2014, p. 1).

To draw a complete picture of how human communication works and what are its origins, linguists have to embrace all sensory and semiotic modalities involved in language interactions. While it is by all means true that, recently, multimodal perspectives on language and communication have been gaining more and more prominence, the overall focus of the study of language still remains strongly skewed toward written modality (described by Per Linell back in 1982 as the *Written Language Bias* of modern linguistics). Although the importance of written communication in today's digital society should not be underestimated, it is reasonable to assume that face-to-face language interaction is the primary mode of human communication (moreover, in some cultures, it is the only mode of linguistic communication).

This study deals with an eminent property of spoken language in everyday use: the intricate interweaving of sensory and semiotic modalities, with a particular focus on the interaction between the audio-oral (speech) and the visuo-motoric (co-speech gesture) modalities. Linguists who study multimodality are inclined to say that gestures *accompany* speech, yet sometimes gestures seem to be something more than a mere accompaniment.

Let us consider the following examples:

(1) *I need a screw about this size.*

(2) *Ann took her housemate's pot without asking.* (Holler and Beattie, 2003)

(3) *"'Good news or bad news?' he said.*
'Uh… bad.'
'No, you get the good news first.' He shook his left hand and opened it dramatically, spoke as if he had released a sentence. "The good news is that I have a tremendously intriguing case for you.'
I waited. 'The bad news.' He opened his right hand and slammed it on his desk with genuine anger. 'The bad news, Inspector Borlú, is that it's the same case you're already working on.'" (C. Miéville: *The City and The City*, p. 130.)

The above examples illustrate just some of the ways in which gestures that accompany spoken discourse (*co-speech gestures*) are integrated in *multimodal utterances* and that will be put under scrutiny in this study.

Example (1) represents an utterance that requires a gesture. In this case, gesture provides an information critical for a proper understanding of the utterances. It is not a part of the syntactic structure of the sentence (that is not deficient without it), but it is an obligatory component of the multimodal construction.

The utterance in the example (2) does not require the presence of gesture. However, since the sentence is ambiguous (because of the polysemous word *pot*), it requires additional context in order to achieve the intended interpretation. From a range of contextual cues that could be available upon hearing this sentence in discourse, gesture may easily prime the target sense of the word pot (by, for instance, enactment of holding the handle of a cooking pot). The gestural information does not have to necessarily be the critical part of the context, but it may be a salient one. When the interpretation of an utternace involves alternative *construals* (Langacker, 1987a), gesture may aid in the construal process by introducing additional perceptional cues needed for the discrimination between the alternatives.

Finally, example (3) is a description of gesture used to emphasize some aspects of the spoken discourse. Such a use of gesture may seem only "ornamental", without a special relevance for the utterance itself, apart from, perhaps, emphasis.

In this particular example, taken from literary fiction, we are presented with an apt description of a the use of co-speech gestures embodying what is called gestural metaphoricity (Cienki and Müller, 2008) based on two simultaneous processes: first, we view the gestures as embodied representations of conceptual metaphors based on objectification of the mental entities, such as IDEAS ARE OBJECTS, THOUGHT IS OBJECT MANIPULATION or COMMUNICATION IS OBJECT TRANSFER (Lakoff and Johnson, 1980). At the same time, the gestural pattern exhibits an iconic mapping between two opposing concepts and their embodied representation in gestures (Wilcox, 2018).

The above examples demonstrate that the semantic integration between gesture and speech occurs with a varying degree of conventionalization and obligatoriness of the presence of gestural component. In this respects, three notions related to the integration of gesture with speech into a single meaningful unit will be central to this study. First is *multimodal utterance* or *multimodal expression* (Enfield, 2009; Kendon, 2004), by which I mean an aggregate of speech and gesture segments combined to convey the same content. A type of multimodal utterance, *multimodal construction* (Andrén, 2010; Zima and Bergs, 2017) is a conventionalized and entrenched multimodal form-meaning pairing that is, in both production and perception, processed as a gestalt-like unit. The third key term is *multimodal construal*, which refers to a cognitive operation underlying the production and comprehension of multimodal utterances.

The question as to under what conditions – if any – a multimodal expression qualifies to be called a multimodal construction belongs among the cardinal questions

of the current *gesture studies* – the interdisciplinary subfield of language sciences focused on the study of gestures and, among other things, their role in language processing, cognition and language development – both ontogenetic and phylogenetic.

Some constructions can certainly be categorised as multimodal. This is the case of, e.g. the constructions with a deictic expression that refers to a gesture (example 1), or ritualized multimodal utterances such as the *sign of the cross* gesture performed together with the Trinitarian formula (*in nomine patris et filii et spiritus sancti*).

This study explores the domain where the constructional status of multimodal utterances is uncertain and fleeting, an area that is largely unexplored. It is concerned with the situations in which co-speech gestures appear to be "tuned in" to speech, emerging from the unwitting stream of gesticulation as profiled elements.

Gesture-speech integration occurs in countless ways and at numerous levels: iconic gestural representation of the semantics of accompanied speech (as in the example 2), iconic representation of figurative conceptual structures (such as example 3), or alignment between gestures and acoustic qualities of speech (gestural peaks are synchronised with intonation peaks), for instance. Iconicity, as a general underlying principle of many types of gestural representations, will be in focus of the present study. The iconic mappings that will be in focus here may not be apparent at first sight, but emerge upon a closer inspection as systematic patterning in the gestural form that may be associated with certain aspects of the content carried by speech. As in the example 3, the kind of iconic mapping in question is not constituted directly between the meaning of a word and the form of a gesture, but is established metaphorically. In particular, I focus on how gestures are associated with *eventuality*: the multifaceted domain of how humans segment their experience into *events* (Kurby and Zacks, 2008; Radvansky and Zacks, 2014), handle them as conceptual objects and how they understand their unfolding in time. Concentrating on selected aspects of the multimodal construals of events (particularly the construals of event boundaries), I will look at how gestures may give away the embodied aspects of our understanding and perception of events.

*

* *

1.1 Objective and some methodological remarks

This study contributes to the exploration of the interface between co-speech gestures and the structure of accompanied speech. Namely, it focuses on the relation between formal characteristics of gestural movement and the linguistic expression of event structure in two languages: English and Czech. These two languages, representing

(West) Germanic and (West) Slavic genera of the Indo-European family, have never been approached comparatively in this respect. Neither have they been systematically compared in terms of event structure expression, nor co-speech gestures alone.

The study was carried out in two stages: first, a corpus survey was performed to analyse the multimodal eventuality expressions spontaneously produced by speakers of English and Czech in interactional settings. Subsequently a behavioural experiment was conducted, designed to validate the findings of the production part of the study from the comprehension perspective (with Czech subjects only).

Multimodal expression of eventuality was addressed previously by a number studies (Becker et al., 2011; Duncan, 2002; Parrill et al., 2013; Cienki and Iriskhanova, 2018, or Hinnell, 2018) and this study builds upon them theoretically and methodologically. However, the approach that is taken here is novel in several respects. First, it puts emphasis on ecological validity, analysing spontaneously produced multimodal utterances in naturalistic settings. Second, it is set in the framework of Multimodal Construction Grammar, an emergent cognitive and usage-based approach (Zima and Bergs, 2017; Schoonjans, 2017), which, among other things, involves a comprehensive, inductive process of data generation, accounting for a variety of factors. Third, it introduces a new way of analysing multimodal data, inspired by corpus-based approaches and particularly suitable for dealing with the specific constraints of multimodal studies.

1.1.1 Transcription, notation and linguistic examples

Throughout this text, the word *gesture* is used both as a countable noun referring to a single instance of gesticulation and an uncountable noun that refers to the phenomenon of speech-accompanying gesticulation.

Interlinear glossing of the linguistic examples as well as abbreviations of linguistic features follow the Leipzig glossing rules (Comrie et al., 2008).

Examples that involve interactions between multiple speakers are given as transcriptions, based on the notation conventions of Conversation Analysis (Jefferson, 2004; see section 3.1).

All languages henceforth referred to in this work are listed in Appendix A (together with a genetic classification).

Throughout this text, small capital letters are used to indicate CONCEPTS or conceptual categories as is customary in the cognitively oriented approaches. Capital letters are used to indicate the English glosses for SIGNS of sign languages. Labels for grammatical constructions are given in brackets in the following way: [*component* + *component*].

1.1.2 Data storage and accessibility

I tried to comply with the commitment to open science as much as possible. For the experimental part of this study, complete documentation is freely available, including stimulus materials, script of the experiment, datasets and R scripts. For the corpus study, annotation files, datasets and R scripts are openly accessible. The Czech part of the corpus itself, however, has not been made public. This is due to a limited licence granted by the subjects that were recorded. The possibility of a free unlimited access to the recordings was excluded prior to the recordings (and the informed consent form was designed accordingly, see Appendix B) because of the character of recording sessions. In order to achieve the highest possible level of naturalness and spontaneity, authentic business meetings were recorded where sensitive or confidential topics were discussed. The recordings thus contain sensitive information including names and confidential content to such an extent that any anonymisation attempt would be futile. At the expense of full data transparency, subjects' personal comfort in this matter helped to achieve the desired degree of ecological validity. The English subcorpus was sampled from the AMI corpus that is freely available under CC BY 4.0 licence on the corpus website: http://groups.inf.ed.ac.uk/ami/corpus/.

All available documentation can be accessed via Open Science Framework here: http://osf.io/ajc29/.

1.1.3 Software

This thesis was typeset in *LaTeX* using *Overleaf*, an on-line writing and publishing tool.[1] I adapted a template created by Martin Jareš, Arnošt Komárek and Michal Kulich (Faculty of Mathematics and Physics, Charles University). Bibliography was organized with the help of *Zotero*.[2] Software tools used for multimodal annotation, statistical analysis and stimulus presentation are credited in the respective sections. Software scripts are not attached in the form of appendices but can be accessed at the OSF repository.

1.2 Outline

Chapter 2 provides an overview of the fundamental concepts related the linguistic study of gesture. A special attention is paid to the cognitive theories of gestural representation. Departing from the traditional analytic concepts of gesture studies, this chapter offers a perspective of co-speech gesture based on three central assumptions: (i) iconicity is an organizing principle of gestural representation across all types of gesture; (ii) every aspect of multimodal representation shares should be explicable within

[1] http://overleaf.com
[2] http://zotero.org

a single general cognitive model; (iii) gestures are inherently multifunctional.

In Chapter 3, I review two theoretical streams within gestures studies: interactionist and cognitivist (with a particular focus on the cognitive approaches). I subscribe to the recent calls for a synthesis between the two approaches, suggesting Multimodal Construction Grammar as a suitable methodological framework that could accommodate the specific requirements of the both perspectives.

Chapter 4 concludes the theoretical part of the thesis by delimiting the titular notion of eventuality and reviewing the empirical research on the role of gesture in multimodal expression of event structure. In this chapter, I present the cognitive model of event semantics and the way it will be applied in the multimodal research. The approach adopted in the present study is based on the notion of *aspectual contours*, operationalised in this study in terms of associations between *aspectual types* of predicates, and a set of formal features of gestures that accompany them.

Chapter 5 presents the design, methods and results of the corpus-based study of Czech and English multimodal construals of eventuality. In the quantitative analysis of the multimodal corpus data, an approach based on recursive partitioning is adopted for the first time in the study of co-speech gestures.

Chapter 6 reports on the experimental study: its theoretical prerequisites, design, methods and results. A gesture perception experiment was ran with the speakers of Czech to validate the results of the Czech part of the corpus study from the comprehension perspective. The experimental design was assisted by a corpus survey, motion-capture data and a lexical rating study.

Chapter 7 summarises the findings of the corpus-based and experimental study. Typological motivation of the observed cross-linguistic differences and the theoretical and methodological implications for Multimodal Construction Grammar are discussed. The chapter is concluded by the customary prospects for the future research.

1.2.1 Author's note

Parts of this text have been published elsewhere in some form. Passages contained in the first three chapters appear (in a more or less adapted manner) in two articles that I co-authored with Eva Lehečková (2018; 2019). Parts of Chapters 4, 5 and 7 are based on the paper Jehlička and Lehečková (2020).

2. Gesture-speech interface

> *"There are some topics covered in every grammar, and other topics that are rarely, if ever, included. One topic likely to be in the latter category is gesture […] From one perspective, this omission makes sense. After all, gesture is not part of the language proper. (Or is it?) But from another perspective, omitting gesture is puzzling simply because wherever people use language – any language – they use gesture too"* (Abner et al., 2015, p. 437).

In spoken language interaction, humans rely on a broad repertoire of meaning-making tools, not only on speech itself. In other words, speakers exploit various semiotic modalities of communication to convey information efficiently . Such combination of different modalities during on-line language production, particularly combining speech with bodily movements (manual, facial, head, shoulder or trunk gestures), surprisingly does not require much effort from speakers. On the contrary, simultaneous multiplicity of modalities seems to be the most natural form language production takes during spontaneous everyday language interaction. Experimental evidence suggests that speakers tend to make use of any modality available, even at the cost of apparent redundancy (cf. the evidence from bimodal bilinguals who employ simultaneous blending of spoken and signed production – see Emmorey et al., 2008). What seems to be contradictory to the principle of language efficiency or parsimony (Martinet, 1960), might in fact be the most effective in terms of cognitive processes related to language production and comprehension.

From the variety of semiotic means that may be combined into a single multimodal communicative channel, it certainly is *gesture* that is visually most salient, ubiquitously accompanying spoken production, even when there is nobody to perceive the production in the visuo-motoric modality.[3] While *gesture*, as it is treated in gesture studies, is a very broad aggregate of different kinds of bodily behaviour, including head movement, eyebrow movement, eye gaze, trunk positioning, posture, foot movement or facial expressions, I will only focus here on manual gestures that accompany the production of spoken language: the so-called *co-speech gestures*.[4]

In this chapter, I will introduce the theoretical basis of how co-speech gestures are generally dealt with in this study, departing from what can be called a standard approach in today's gesture studies (Section 2.1. The standard categorization of co-speech gestures will be revisited in the light of current cognitive theories of *iconicity* (Section 2.2.

Although the relation of gesture and language has various facets, one particular aspect has a prominent position in gesture studies: namely gesture accompanying

[3] A typical example: when speaking on the phone.

[4] Henceforth, I will use the term *gesture* in reference to co-speech gestures.

spoken production of language, the omnipresent gesticulation that we carry out unwittingly, effortlessly and constantly while talking to each other. Various kinds of gestural behaviour have been classified according to the perspective from which gesture is approached (e.g. Argyle, 1975; Birdwhistell, 1970; Ekman and Friesen, 1969 among many others). I shall not list them all, since one approach to gesture classification has been dominant in gesture studies and is generally agreed upon as a standard model among the majority of gesture researchers. A detailed discussion of various classification systems can be found in Kendon (2004) or Bohle (2014).

2.1 Continua

Kendon (1988, *inter alia*) and McNeill (1992) argued for a scalar categorization based on the linguistic nature of gesture. According to this view, there is a scale on which *gesticulation* – seemingly meaningless and involuntary movements of hands accompanying speech – occupies one end, whereas *signs* – lexical items of sign languages – occupy the other. In between, there are also *emblems* (conventional, culture-specific gestures such as 'thumbs up'), *pantomime* (enactment of actions), and *speech-related* gestures (the primary focus of gesture studies) in varying order according to their specific linguistic aspects. This scale was proposed by David McNeill (1992), who called it *Kendon's continuum*[5] and two parts of the four-fold scale are depicted in two Figures below (Figures 1 and 2)[6]. From a linguistic point of view, the most prominent category of gesture is represented by gesticulation, i.e. gestures produced along with spoken language, (typically) involuntary and omnipresent across languages (but not universal as for their formal properties). In gesture studies, the term *co-speech gestures*[7] has prevailed over *gesticulation* and is now widely used by linguists and other social scientists studying non-verbal behaviour.

⟶	⟶	⟶	⟶
GESTICULATION	EMBLEMS	PANTOMIME	SIGNS
obligatory presence of speech	optional presence of speech	obligatory absence of speech	

Figure 1: *Continuum 1: relation to speech*

Whereas there is hardly any doubt about the continuum regarding the co-occurrence of gesture and spoken language (Figure 1), the scale of the presence or absence of linguistic properties (Figure 2) is not straightforward. The degree of presence of linguistic

[5] Adam Kendon himself, however, did not explicitly posited any kind of continuum concerning communicative gestures and expressed reservations about it (personal communication between A. Kendon and J. Fulka).

[6] The Figures 1–4 are taken from McNeill, 2005, p.7–10.

[7] Coined by Kendon, 1994

⟶	⟶	⟶	⟶
GESTICULATION	PANTOMIME	EMBLEMS	SIGNS
linguistics properties absent		some linguistic properties present	linguistic properties present

Figure 2: *Continuum 2: relation to linguistic properties*

properties here corresponds to the extent to which gesture is constrained by the grammatical system – with respect to the gesture's morphology as well as syntax. Thus, this continuum also reflects the degree of grammaticalization or conventionalization (Figure 3).

⟶	⟶	⟶	⟶
GESTICULATION	PANTOMIME	EMBLEMS	SIGNS
not conventionalized		partly conventionalized	fully conventionalized

Figure 3: *Continuum 3: relation to conventions*

The problem of the presence or absence of linguistic properties becomes more complicated when it comes to the semiotic status of gesture. In terms of Peirce's typology of signs (1931), a gesture may become an icon, an index as well as a symbol based on the relation to the concept it refers to. Although this certainly applies to sign language signs, which are primarily defined as arbitrary symbols, the role of iconicity in sign language systems is undeniably essential (Klima and Bellugi, 1979). Indexicality also plays an important role in the pronominal system across many sign languages). The same applies to emblems, whereas pantomime's primary semiotic status is iconic, though not exclusively. The semiotic status of gesticulation is somewhat obscured by the wide range of functions it may have in communication. I will elaborate on this in more detail below.

⟶	⟶	⟶	⟶
GESTICULATION	PANTOMIME	EMBLEMS	SIGNS
global global & synthetic	global & analytic	segmented & synthetic	segmented & analytic

Figure 4: *Continuum 4: character of semiosis*

Finally, the fourth dimension of the continuum ("character of semiosis") captures (Figure 4) how "language-like" various types of gestures are in the way they carry meaning. This dimension entails two axes: compositionality (*global* vs. *segmented*) and mapping between semantic and formal units (*synthetic* vs. *analytic*). Let us focus solely on the extreme ends of the scale. According to McNeill, gesticulation is "global and synthetic" as opposed to sign language (or language in general, for that matter), which is "segmented and analytic". Gesticulation is non-compositional (i.e. global): "The meanings of the 'parts' of the gesture are determined by the meaning of the whole" (McNeill, 2005, p.

10). Sign language utterances, on the other hand , are considered compositional – segmentable into lexical units – in the same way spoken sentences are segmentable into words. Synthesis here refers to the fact that a gesture can accumulate, in a simultaneous manner, multiple "meanings" that are distributed "across the entire surface of the accompanying sentence" (*ibid.*). Sign languages are, in contrast, considered analytic.

This dimension is perhaps the most problematic of the four. We will see (in Section 3.2.3) that non-compositionality and idiomaticity are defining features of every grammatical construction – from lexical items to complex discourse structures. This of course applies to sign language utterances too. As for the synthetic-analytic distinction, distinguishing between gesture, spoken language, and sign languages in these terms does not make much sense either. First, such a view is Anglocentric: spoken languages of course differ in how semantics is encoded in morphology, the above claim about the accumulation of meanings into a single gesture is also true for lexical units of polysynthetic languages (e.g. Yup'ik). Sign languages, on the other hand, if viewed through the lens of spoken language morphological typology, have been likened to synthetic types rather than analytic (Klima and Bellugi, 1979). This is due to the key property of sign language signs: the ability to simultaneously convey multiple semantic features, mapped on the co -articulated "morphemes". McNeill argues that gestures do not contain morpheme-like units, since a specific formal feature (e.g. type of movement of a handshape) may take up a different meaning each time (McNeill, 2005, p. 11). That indeed captures a critical difference between gesture and sign languages, but this difference does not really involve the character of semiosis – it is again only a matter of conventionalization. McNeill's point is to highlight the different modes of representation inherent to gestural and verbal expression – this is the foundation of his Growth Point Theory, which will be discussed in detail in Section 3.2.2. The opposition of gesture and sign language is not, in fact, entirely relevant to McNeill's argumentation. In the fourth dimension of the continuum, it appears to be misconstrued. Historically, focusing on the differences between sign language and co-speech gestures (particularly on how sign language is pervasively superior to co-speech gesture) had its justification in the ideological struggle for emancipation of sign linguistics (the same way as there was a tendency to sideline the iconic properties in sign languages, e.g. Newport and Supalla, 1980). Bearing in mind the fundamental differences between sign language and gesture, (such as the degree of conventionalization mentioned above, or the fact that the cognitive processing of sign and spoken languages shares the same neurological mechanisms – which is not the case of gesture), exploration of the interfaces between signs and gestures provides important insights (see, e.g., Volterra et al., 2017, or Ortega and Özyürek, 2019).[8]

[8]This will be illustrated by presenting relevant evidence for the Event visibility hypothesis discussed in Chapters 4.

2.2 From gesture types to dimensions

Co-speech gestures apparently represent a very diverse class when it comes to their functions in communication. From the variety of classifications (some of which can be found in the references at the beginning of this section), one has become a standard part of the descriptive apparatus of today's gesture studies. The standard typology operates with four types: *iconic* gestures, *metaphoric* gestures, *beat* gestures and *deictic* gestures. The first three types were first identified by David McNeill and Elena Levy (1982)[9] in their pioneering gesture analysis (I will address this study in more detail in Chapter 4). These four types represent a rather diverse assemblage of phenomena; we can see that the types are not singled out based on a single criterion but can be viewed as a combination of semiotic and functional aspects of the gesture-speech relationship. Iconic and deictic categories have a clear semiotic motivation, echoing Peircean categories of *icons* and *indices* respectively.[10] Beats, on the other hand, cannot be defined simply in semiotic terms; rather, they represent a category specified by function (parsing of the speech stream into rhythmic segments).

Semiotic or functional criteria were not, in fact, the central point of McNeill and Levy's classification. The authors of the original triad themselves strove to capture the degree of "correlation" between gestures and meaning of the accompanied speech. It is important to note that McNeill and Levy's categories are rooted in an early cognitive theory, according to which gestures were assumed to be visible manifestations of conceptual representations (p. 272). The original typology, although involving three types, is, in fact, binary, the main defining feature being the level of representation. Since the beginning of gesture studies, there has been a tendency to draw a division line between representational gestures and beats. This is one of the common and entrenched misconceptions concerning gestures, in part caused by McNeill's own writings, in which the beat gestures were sometimes treated inconsistently in this regard (cf. McNeill and Levy, 1982; McNeill, 1992, vs. McNeill, 2005).

In their 1982 paper, McNeill and Levy certainly did not exclude beats from representational gestures. On the contrary, they explicitly acknowledged that beat gestures are also manifestations of conceptual representations. But unlike iconics and metaphorics, beats are bound to a higher level of representation, not at the level corresponding to individual words or low-order syntactic structures, but at the level of discourse organization. This is a crucial observation, and the fact that it has been neglected by many gesture researchers has had a considerable impact on the development of the field (e.g. the psycholinguistic theories discussed in Section 3.2.2).

Another common misconception concerning gesture types that will be countered here is their presumed discreteness. The four types are indeed often treated as

[9] As "iconix", "metaphorix" and beats respectively.

[10] Metaphoric gestures are, from the Peircean perspective, treated as icons. In Peirce's semiotic theory, metaphors are a subtype of icons (see below).

natural categories – as if an individual gesture should fit in a single pigeonhole. Even though McNeill and Levy coded their gestures in this way in their 1982 study, their choice was justified for several reasons. As one of the first attempts at a systematic description of the gesture-speech interface using quantitative methods, the analysis was based on a manual coding of gestures which was carried out in a rather simplistic manner, using a material that allowed for it: elicited narratives with an abundance of "depictive" gestures. McNeill himself later made clear that the types (by then expanded into a "quartet" including deictic gestures) were not intended as discrete and mutually exclusive categories but rather dimensions that may be traced in naturally occurring gestures simultaneously in various configurations (McNeill, 2005).

Below, I review the four semiotic-functional dimensions, understood here as ideal, constructed categories, not natural ones. First, I will recapitulate the definitions provided by McNeill and Levy for the gesture types that correspond to the respective dimensions; then I will focus on the recent shift beyond the original classification, towards a new perspective stemming from research on the role of iconicity in multimodal communication. For each dimension, a cognitive account will be suggested.

One of the aims of the following sections is to illustrate that the four types defy a discrete, categorical conception, as they simply cannot be treated separately. We will see that the notion of *iconicity* runs like a scarlet thread through all dimensions, providing a key to a unified approach to gestural representation.

Iconicity

It would not be sheer speculation to assume that most attention has been paid to iconic aspects of gestures and that iconic gestures are by far the most represented gestural type in gesture studies. First defined by McNeill and Levy as "a formed gesture which depicts in its form or manner of execution aspects of the event or situation being described verbally" (McNeill and Levy, 1982, p. 275)), an iconic gesture may seem as a quite easily recognizable phenomenon (at least in theory). An example of a clearly iconic relationship between gesture form and the meaning of its verbal counterpart may be a repeated circular movement of one's index finger depicting a cyclic movement while describing a wheel spinning (see Figure 5 and example 4).

Modes of representation

Even in this seemingly straightforward gesture, we can identify at least three parallel processes in which the form-meaning mapping is established. At the most basic level, the hand represents the wheel – or in this case a part of the wheel, the tip of the finger corresponding to the wheel's rim. Since the fingertip alone cannot stand for anything else than a certain point of the curve of the wheel's edge, there must also be the movement of the finger that makes the wheel's shape circular. Then, as the movement

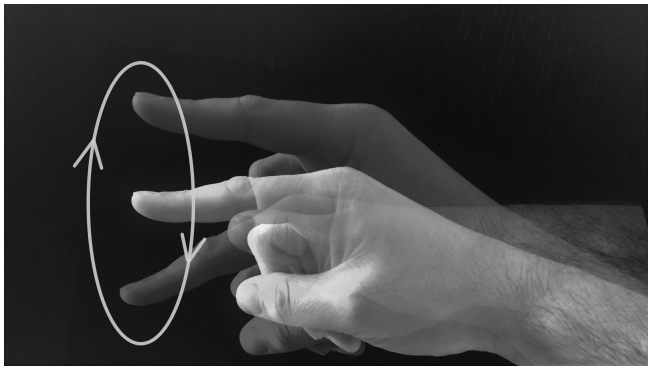


Figure 5: *A cyclic gesture*

(4) *The wheel is spinning.*

continues, the gesture becomes a *representamen*[11] not only of the wheel itself, but also of its moving (rotation around its axis – should the hand remain at the same position). We can thus identify the three processes of *mimesis* that come into play here: (1) the hand stands for the object, (2) the finger traces the shape of the object, and (3) the rotation of the hand enacts the movement of the object. The three processes constitute a non-compositional unity (or a gestalt), as one cannot say that, for instance, the tracing phase is over when the finger completes the first full circle, immediately followed by the enactment, as even a static circular object may be depicted through repeated circular movement. Of course these are not all the possible ways iconicity can be exhibited in gestural representation. Cornelia Müller (1998) distinguished four modes of representation in iconic gestures: *drawing*, *molding*, *acting*, and *representing*. As drawing, molding and representing could be all subsumed under the umbrella of depiction,[12] the classification of iconic representation in gestures may thus be generalized using just the dichotomy of *depiction* on the one hand and *enactment* on the other.

Crucially, these two general mimetic processes do not constitute a binary opposition – both processes may be simultaneously involved, typically in gestures of tool manipulation.[13, 14]

Following in a similar vein, we may add several other dichotomic axes as conve-

[11] In this section, iconicity is approached through the theoretical framework of Charles Sanders Peirce's semiotics (here cited from *Collected Papers* [CP], Peirce, 1932). Iconicity (or any other process of signification) is here understood as a *triadic* relationship between *representamen* (realization of the sign, in case of iconic gestures, the gesture itself), *object* (referent) and *interpretans* (concept) (CP 2.228). Peircean account of iconicity in gesture is provided by Mittelberg (2014; Mittelberg and Evola 2014).

[12] I do not classify representing as enactment, but rather as embodied depiction.

[13] See the "car wreck" example below.

[14] Cf. also the issue of noun incorporation in sign languages (Meir, 2001).

nient ways of carving up a complex concept such as mimesis in gesture. The first axis considers the dynamics of the representamen. Representation of static entities would tend to be subject to depiction, whereas representation of dynamic actions would be achieved via enactment (rather than, say, a series of static depictions). Thus we can distinguish *object-centred* and *action-centred* representations according to the nature of the object (in the Peircean sense). This is important to bear in mind in order not to confuse the object- vs. action-centring with the cognitive process of profiling (addressed below).

The second axis I consider to be a useful descriptive tool in the morphology of iconic representation in gesture is related to the *embodiment* of the mimetic process (not the embodiment of the conceptualization itself). Along this axis, we can distinguish *embodied* representations, in which the vehicle of mimesis is a body part – as in Müller's representations, an open hand, for example, may represent a table. *Exbodied* representations are typically depictions – imaginary scenes created by gestures of drawing or molding – such as tracing the shape of an object with one's fingers. The distinction between embodied and exbodied representation results from the nature of the *material carrier*,[15] not from the generally embodied nature of conceptualization.[16] The difference as it is understood here could be explained from the perspective of perception of the gesture: exbodied representation can be interpreted via the imaginary trace[17] the articulator (e.g. the hand) leaves in the air, upon the desk etc., as opposed to embodied representation, which only uses the body as the carrier. Again, in real-world situations, both modes may blend: consider a gesture of drawing (exbodied virtual depiction) with fingers on the palm (embodied representation of a sketchpad).

The focus on the object or action corresponds (partly) to what McNeill calls gestural viewpoints (2005). McNeill observed two perspectives gestures may take. In narratives (re-telling of a cartoon the gesturers have just seen), the speakers produced their representational gestures either from what McNeill called the *observer viewpoint* (OVPT), i.e. "third person" representation of entities (objects, persons, abstract concepts) expressed in speech, or "first person" *character viewpoint* (CVPT), in which the speakerp lays the part of a character and produces gestures by representing the person by acting. However, the character viewpoint does not apply to animate characters only, but to inanimate entities that afford "personification" or embodied enaction as well.

There is not a complete overlap between the viewpoint configurations and object- or action-centring. For instance, in the case of bumping one's fists together representing a collision of two cars, the perspective would be OVPT, but the gesture itself would be a blend of object-centred representing (fist as a car) and action-centred acting

[15] "[T]he embodiment of meaning in a concrete enactment or material experience" (McNeill 2005, p. 98, cf. also Vygotsky, 1986).

[16] Cf. also Mittelberg's unrelated usage of the term exbodiment (2013).

[17] This corresponds to what Mandel (1977) calls "virtual depiction" in the context of iconic representation in American Sign Language (ASL).

	depiction	enactment	embodiment	object-centred	action-centred
drawing	+	-	-	+	-
molding	+	-	+/-	+	+/-
acting	-	+	+	-	+
representing	+	-	+	+	-

Table 1: *Modes of representation*

(bumping as crashing).

Table 1 shows that the aforementioned features of iconic representation do not overlap with Müller's four categories, categories, which may serve as a starting point for a more fine-grained description of the mimetic processes involved in a particular instance of gestural representation.

This is of course only a portion of the possible aspects of mimetic processes coming into play while constituting an iconic mapping between the object and the representamen. However, as is evident from the examples above, the mapping is not always so straightforward and relatively easily deconstructable as in the case of e.g. one phonological[18] feature (e.g. a finger) representing a single entity.

By the late 2010s, the topic of iconicity in language emerged from the obscurity where it had been cast off for most of the history of modern linguistics, and has since become one of the key issues taken up by many researchers to fill the long-standing niche across subfields of linguistics – acquisition (Perniss et al., 2018; Laing, 2019), language evolution (Perlman, 2017) or neurolinguistics (Vigliocco et al., 2020). One of the impulses to revisit iconicity in language, as is often the case, came from sign linguistics. After it had shaken off its historical disregard of the motivatedness of signs, it embraced iconicity as a fascinating property of language in general, which lead to a growing number of studies on iconic devices in various sign languages (Perniss, 2007, on German Sign Language (DGS), Arik, 2012, on Turkish Sign Language (TID) or Börstell and Östling, 2017, on Swedish Sign Language (SSL), and many others), eventually resulting in synthetic accounts of iconicity as a crucial aspect of communication across modalities (Perniss et al., 2010; Perniss and Vigliocco, 2014).

For gesture studies, the main boost provided by the renaissance of iconicity research was not acknowledging the role of iconicity – it has always been a prominent topic inside the field – but rather abandoning the simplistic[19] view of this phenomenon and moving forward towards a more sophisticated view of iconicity reflecting (1) the diversity of types of iconic mappings, (2) different structural levels at which these mapping may function, (3) scalar nature of iconicity, and (4) how iconicity is constrained by general cognitive processes related to perception or memory. Let us briefly review

[18] As in sign language phonology.

[19] Clear cases – one-dimensional iconic gestures, exemplified by cases observed in controlled environments produced typically by speakers retelling the same cartoon storyline.

how taking these four aspects into account enhanced our understanding of gestural iconicity.

Types of iconicity

For the traditional typology of iconic mappings, we may again refer to Peirce's theory. Peirce distinguished three kinds of icons: *images* – representamina sharing physical qualities with the object, *diagrams* – representamina sharing internal structure with the object or representing the relationship between several objects, and *metaphors* – "[representamina] which represent the representative character of a representamen by representing a parallelism in something else [compared to images or diagrams]" (CP 2.277). Setting metaphors aside for now, we are left with two types, usually labelled *imagistic* (sometimes also as imagic) and *diagrammatic* iconicity.

The imagistic type is what first comes to mind with iconicity, the examples being, onomatopoeic words, pictographs or iconic gestures (such as the above examples). In diagrammatic iconicity, it is not the physical appearance of the object itself (in the first place) which is subject to mimesis, but the internal structure of the object. A telling example of a diagram is the "Olympic rings" logo – the five coloured rings do not represent the continents by themselves (individually, they are *symbols*)[20] but combined they resemble the parts of the whole – they copy the morphology of the object (in this case the Earth's landmass). Mittelberg (2008) reports several examples of diagrammatic iconicity in gestures accompanying discourse on grammatical relations. For instance, the syntactic structure of a sentence, for instance , may be diagrammatically mapped onto a series of gestures in which the speaker visualizes the boundaries of syntactic units using both of their hands, placing the hands along a horizontal axis, as in a linear representation of constituent structure using brackets. In this particular case, an isomorphic relation is being established between representamen and the structure of an object in a metaphoric way – syntactic structure is not a real-world object with any physical qualities. Mittelberg points out that the three types of iconic mapping cannot be treated as exclusive categories but may in fact simultaneously partake in representational processes in gesture.

Dingemanse (2011) elaborates on the diagrammatic type, breaking it down into two subtypes. First is gestalt iconicity, defined in the context of lexical forms as "a type of diagrammatic iconicity in that a relation between forms (the parts of the word) has a resemblance to a relation between meanings" (Dingemanse, 2011, p. 167). At the morphological level, this type of iconic mapping exhibits itself clearly e.g. in morpheme reduplication where the strings of morphemes correspond to the character of an event or object. But can we identify such iconic mapping in gesture? According to the fourth

[20]Contrary to popular belief, neither are the colours of the rings iconic representations of the respective continents – they represent major colours represented in flags worldwide (via *pars pro toto* metonymy). That, in fact, constitutes a "diagram within a diagram".

dimension of Kendon's/McNeill's continuum (Figure 4), gesticulation is "global and synthetic" – therefore we should not be able to segment gestures into morpheme-like units (only to "phoneme-like" parameters such as handshape or movement manner as in sign language lexical units). Yet, we can indeed find cases of gestalt iconicity in gesture. Consider the following example (5, Figure 6) from a Czech multimodal corpus (see Chapter 5 for the details on the source).

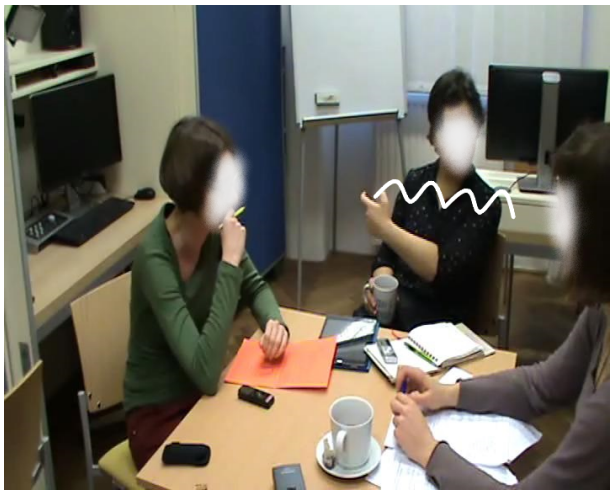


Figure 6: *Example 9: Gestalt iconicity*.

(5) *pak* *bych-om* *měli* *roz-fáz-ovan-ou* *tu* *praxi*
then be.COND-1PL have-PST.1PL DISTR-phase-PTCP-ACC.F DEM.ACC.F training.ACC

'Then we would have the training divided up into phases.'

The gesture that the speaker produces together with the word *rozfázovanou* ("divided-up into phases") is a complex gesture phrase (see below) consisting of multiple strokes that iconically represent the internal structure of the event expressed verbally not only by stroke repetition (exactly as in morpheme reduplication) but also by the "slicing" gesture based on a metaphor of cutting time into units (see also Calbris, 2003, or Tversky and Jamalian, 2012).

The second subtype, *relative* iconicity, on the other hand

> "involves mapping a relation between forms onto a relation between meanings. Like Gestalt iconicity, it is a type of diagrammatic iconicity; but unlike Gestalt iconicity, which focuses on the internal structure of signs and meanings, relative iconicity concerns a relation between multiple signs that has a resemblance to the relation between multiple meanings" (Dingemanse, 2011, p. 170).

Dingemanse illustrates this subtype on Siwu *ideophones* – words that express complex (multi-)sensory experiences employing iconic devices such as sound symbolism. In

case of relative iconicity the mapping is constituted between relations and is achieved by mapping, e.g. the relation between the sound qualities of two phonemes mapped onto a relation of another kind via sound symbolic properties of the individual phonemes (or, in this case, by sound symbolism related to a phonetic feature in which the two phonemes differ). In many languages of Western Africa sound symbolic correspondences play a productive role in word formation, with specific phonetic features associated with clusters of semantic features of physical and abstract qualities. In an early account of sound symbolism in the lexicon, Westermann (1937), for example, presents a crosslinguistic analysis showing correspondences between the back and front vowel opposition (combined also with low vs. high tones) and dichotomies such as WEAK VS. STRONG, THICK VS. THIN OR ROUND VS. STRAIGHT (in the respective order).

Gestural relative iconicity is sometimes closely intertwined with sound-based metaphorical mappings similar to those underlying the sound symbolic word classes. When talking about music, speakers of many languages rely on a spatial metaphor of pitch (low frequency tones are low, high frequency are high).[21] The staff notation is framed by the spatial metaphor of pitch, and at the same time, it is given as an example of diagrammatic iconicity. Blending the spatial metaphor of music with the conceptual metaphor UP IS GOOD, DOWN IS BAD (Lakoff and Johnson, 1980), then becomes an expressive tool of musical composition. The contrast between high and low notes and positive and negative concepts is one of the elementary meaning-making devices in music.[22] Musical metaphor has, for quite a long time, been of interest to analyses inspired by cognitive linguistics (Zbikowski, 2002). One of the great examples of how the crossmodal conceptual metaphor works in a musical composition is Richard Wagner's opera *Tristan und Isolde* from 1865. The entire work is built upon the tension between two emotions – longing (*Leiden*) and desire (*Sehnsucht*) which are marked by rising and falling melodic motives respectively. The famous "Tristan chord", introduced in the overture and then reappearing throughout the score, combines the two motives in a striking dissonance, representing the tension between the two emotional states (Figure 7).

Opera is, of course, a multimodal art form – the full expressive force of music needs to be complemented not only by the acoustic component, but also by the singers' acting, an important aspect of which is gesture. In the so-called Bayreuth style of staging Wagner's operas, iconic gestural co-expression of the vertical metaphor central to the two motives was a preferred way of acting in this particular piece (Baragwanath, 2007). Using the example of the sopranist Anna von Mildenburg, a prominent Wagner actor advocating the Bayreuth style, Nicholas Baragwanath points out that in her performances in Tristan und Isolde "[her] gestures, in contrast [to Wagner's own

[21] The spatial metaphor is not universal – see below

[22] It is of course not the pitch alone that constitutes the meaning of a musical piece, but also colour (timbre), rhythm and other means.

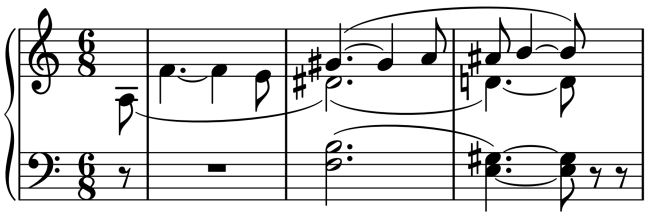


Figure 7: *Tristan chord in the third measure of the prelude to Act I of* Tristan und Isolde *(piano transcription)*

direction notes] mirror the musical contours precisely. The movement of the right arm coincides with the fortissimo high-note at 'dass hell', the collapse onto the steps with descent and diminuendo of the Bliss Motiv (…)" (*ibid.*, p. 70). Such use of gesture produces a truly multimodal co-expressive assemblage of music, libretto and gesture, from which the diagrammatic structure of the (overall metaphoric) complex clearly stands out. The form-meaning mapping is in this case two-fold: the first layer is constituted by a diagrammatic iconic relation between the musical pitch and text of the libretto, the second layer emerges from the diagrammatic relation between the text and the gesture.

Upward and downward hand movements were shown to accompany metadiscourse on music by musicians and laymen alike (Lemaitre et al., 2017) and the multimodal metaphor of vertical movement was also reported to have a positive effect on L2 acquisition of tonal languages such as Mandarin or Japanese – from both production (when used by learners) and comprehension (when used by teachers) perspectives (Morett and Chang, 2015; Kelly et al., 2017). The vertical metaphor of pitch is not a universal phenomenon (Dolscheid et al., 2013; Eitan and Timmers, 2010). In many cultures, different source domains of pitch metaphors are found, such as WEIGHT (Kpelle) or AGE (Suyá), or even dichotomies based on specific concepts like, in Shona language, "crocodile" (low pitch) with "those who follow crocodiles" (high), and "stable (person) who holds the piece together" (low) vs. "mad person" (high), as well as "old men's voices" (low) vs. "young men's voices" (high), "men's voices" vs. "women's voices," and "thin" (low) vs. "thick" (high)[23] (Eitan and Timmers, 2010, p. 406). The contrast between THICK and THIN is also the source domain for pitch metaphor in Turkish and Persian, among other languages. Comparing Swedish and Turkish speakers' production of gestures accompanying description of musical pitch, Christensen and Gullberg (2016) observed language-specific differences: in the Swedish subjects, there was a strong tendency to employ the vertical metaphor in gesture, whereas in the Turkish subjects there was a tendency not to represent the pitch metaphor gesturally.

It is worth noting here that at various levels of sound perception, from the physi-

[23] Cf. the examples of semantic oppositions expressed by sound-symbolic words mentioned above.

ology of the ear to the so-called tonotopic map in the auditory cortex, the spatial distribution of areas dedicated to low and high frequencies provides an "explicit representation of frequency" (Bendor and Wang, 2006). Together with other striking isomorphic mappings in the brain (not limited only to human or even to primate) mentioned by Givón (1991), this might be the key to the physiological grounding of iconicity and its importance in cognition.[24]

Metaphorical representation of contrasts is only one instance of relative iconicity in gesture. In this case, the structure being subject to mimesis is binary (or a two-dimensional scale). The above example of gestures diagrammatically representing syntactic structure (Mittelberg and Waugh, 2009) should also be considered an example of relative iconicity – with more complex structures being represented by gestures.

Degrees of iconicity

The shift from a monolithic view of iconicity towards a more adequate model has also been marked by the acknowledgment of the graded nature of iconicity. Again, sign linguistics has been a spearhead; the scalar view of iconicity has been present and established since the very beginning of the field. Klima and Bellugi (1979) introduced a four-fold classification of signs according to their decodability by hearing non-signers (*translucent* – easily interpretable by non-signers, *transparent* – decodable if hints are provided, *obscure* – perceived as apparently iconic, but undecodable, *opaque* – perceived as apparently arbitrary).

Teasing apart the notions of iconicity and transparency in sign languages, Occhino et al. (2017) argue that iconicity, unlike transparency, does not involve only motivatedness by physical (i.e. visually perceived) qualities of the referent. With a theoretical grounding in cognitive approaches to iconicity (addressed below) and supporting their claim by the results of an experiment in which monolingual signers of ASL and DGS rated the iconicity of native and non-native signs, the authors argue that what the recipients interpret as iconic is in fact not driven by some inherent property of the sign itself. Rather, the recipients' individual knowledge and language experience leads to language-specific (and ultimately subjective) strategies of iconicity construal.

Although the assumptions about perception of iconicity in sign languages cannot be automatically applied to the perception of gestures accompanying spoken languages, a lesson should nonetheless be learned by gesture researchers about a number of issues related not only to the scalarity of iconicity, but also to the context in which the degree (and general character) of iconicity of a given representation is analysed. The focus on isolated units (which is often coerced by the methodological and technical constraints), whether it is on single gestures, lexical units, or a combination of both, is necessarily reductionistic as the natural processing never occurs in isolation.

When it comes to the graded nature of iconicity, a major issue arises as to how

[24] Cf. also the so-called *sensory homunculus* in the human brain.

to capture the degree to which an actual sign (be it a sign language sign, gesture or a word) exhibits iconic features. In psycholinguistics, there is a battery of techniques dedicated to the collection of various lexical norms, used. These norms include an array of related properties such as *imageability* (Paivio et al., 1968), *specificity* (Spreen and Schulz, 1966), or *concreteness* (Gilhooly and Logie, 1980), defined as the degree to which a word invokes a mental image, degree of specificity/genericity of the referent and the degree to which the referent can be experienced by senses, respectively. Typically, participants are asked to rate a list of words on a 5- or 7-point Likert scale.

Iconicity, another kindred concept, had long been neglected in psycholinguistics, though it is a potentially relevant semantic norm. First norms for iconicity emerged in the 2000s, starting with sign languages (Vinson et al., 2008). In spoken languages, Perry et al. (2015) collected iconicity ratings of 592 English and Spanish words, and Winter et al. (2017) replicated the English part of Perry et al.'s data and extended it to 3001 words (various parts of speech), while controlling for other lexical features. These included subjective measures such as imageability or concreteness and objective measures such as systematicity (regularity of form-meaning correspondence represented by a corpus-based systematicity index (Monaghan et al., 2014) and frequency).[25] Winter and his colleagues found that iconicity is strongly correlated with *sensory experience* (Juhasz and Yap, 2013) operationalized as "actual sensation (taste, touch, sight, sound, or smell) [experienced] by reading the word" (p. 161), showing that words related with perception (across sensory modalities) tend to score higher in iconicity ratings.

Unlike lexical units in sign and spoken languages, ready-made norms for gestural iconicity had for long remained a desideratum due to the obvious obstacle: the lack of conventional parameters that would allow for constructing a list of comparable gestural forms. Finally, in 2019, Gerardo Ortega and Aslı Özyürek presented the first collection of iconicity ratings of a balanced set of silent gestures. In order to assess the systematicity of gestural forms, they first elicited gestures from 20 speakers that were presented with 272 words representing concepts from different semantic domains and asked them to produce "silent gesture[s] that conveyed the same meaning as the word[s]" (p. 5). Subsequently, the gestures were analysed as to their systematicity based on their phonological features (handshape, orientation, movement and position), leading to a selection of 109 concepts, for which gestures were produced in the same form by most of the subjects. For the final set, a representative gesture was selected for each concept according to the predominant mode of representation, based on Müller's four modes discussed above plus an additional mode – personification (a type of representing where the referent is a person).[26] Acting was the most common

[25] Iconicity itself was operationalized in terms of participants' self-estimation of their ability to guess the meaning of the target words should they not know English. Ranging from -7 to 7, the rating scale also captured negative iconicity (word sounds like the opposite of its meaning).

[26] It is worth noting that the authors found expectable correspondences between modes of represen-

strategy overall, significantly predominant in the action-related domains.

This set of gestures was subsequently used for the rating study with 19 speakers of Dutch judging on a 7-point scale "how well each gesture represented each concept" (p. 9). The ratings revealed interesting interactions between iconicity and semantic domain, as well as representation strategies. The lowest scores were observed in the category of non-manipulable objects (mostly represented by drawing), while action/manipulation related categories (represented by acting) and the category of animate entities (personification) were generally rated as highly iconic. The authors suggest that such results could be interpreted in support of simulation-based theories of gesture (see discussion in 3.2.1) due to the prevalence of representation based on direct embodiment. This also provides more basis for the above argument about modes of representation falling under two clusters – *enactment* and *depiction*.

As for the iconicity ratings alone, interesting patterns were observed in interactions between strategies and domains, with relative low iconicity scores for objects represented by drawing. This is possible because the drawing strategy allows for less suggestive representations than enactments, which, again, indicates that enactment might have a key role in conceptual representation in visual modality.

Due to the lack or insufficient scope of iconicity ratings for lexical items in many languages (let alone gesture forms), researchers often have to resort to gathering ad hoc ratings. In such cases (including the present study), a balance between the definition of the norm that is being rated and the way it is conveyed in the questionnaire, which should be as straightforward as possible, is crucial.[27] Understanding iconicity as a continuous phenomenon also has consequences for experimental studies – it needs to be treated accordingly in operationalization and analysis of iconicity-related variables (see also Sections 6.1 and 6.2).

Metaphoricity and "cognitive iconicity"

Metaphoric gestures are a subtype or a special case of iconic gestures. According to McNeill, metaphoric gestures are based on the same principle as iconic gestures, but the referent they depict is an abstract concept and not a concrete object.

> "A metaphoric gesture according to our definition iconically depicts the vehicle of a metaphor. Unlike a true iconic gesture, a metaphoric gesture does not directly reproduce its meaning (its tenor, which may be in any case unrepro-

tation and semantic groups: "acting was the preferred mode of representation for actions with objects, actions without objects, and manipulable objects; acting and drawing were the main strategies for non-manipulable objects; and personification was favoured for animate entities" (p. 9).

[27] As we have seen (and will again see in Chapter 6), instructions for the raters may be formulated in very similar ways in the case of imageability, concreteness as well as sensory experience – the differences (which may even be found statistically significant, as in Winter et al., (2017) are in fact often based on tiny nuances in how the norms are operationalized.

ducible), but conveys this meaning indirectly, as in a verbal metaphor, through the vehicle" (McNeill and Levy, 1982, p. 289).

But is such a clear-cut distinction between "true" iconicity and metaphoricity relevant or useful at all to the study of gesture's function in communication? The above section shows that it is almost impossible to talk about iconicity and metaphoricity in gesture separately, as the underlying principles are mostly confounded and in real gesture production, it is often hard to tell where "direct reproduction of meaning" ends and metaphoricity starts.

Apart from the initial impetus by George Lakoff and Mark Johnson's theory of conceptual metaphor (1980, Johnson, 1987, Lakoff, 1987), several other theoretical concepts have framed the cognitive approach not only to metaphoricity but to the nature of conceptual representation in gesture in general.[28] Namely, these were Ronald Langacker's notions of *construal* and profiling introduced in his *Foundations of Cognitive Grammar* (1987a) and also related concepts of Charles Fillmore's *Frame Semantics* (1982), though these are not referred to as frequently as Langacker's concepts in the context of gesture.

The principles of attribution of meaning to grammatical structures put forth by both Fillmore and Langacker are applicable to the processes of gestural representation, assuming that both types of representations are driven and constrained by the same cognitive operations.[29]

Concerning the cognitive foundation of form-meaning mapping emergence, the notion of construal appears to be particularly useful, as it incorporates other key concepts of CL such as *image schemata* or *conceptual metaphor*. In their dynamic approach to conceptualization, William Croft and Alan Cruse (2004) single out four core operations underlying construal processes: (i) *attention/salience*, (ii) *judgement/comparison*, (iii) *perspective/situatedness* and (iv) *constitution/gestalt*. These operations are not specific to language – they rank among the general cognitive capacities. Each of the operations that contribute to the construal of a linguistic meaning may likewise be applied to describe how gestural representation works. One of the central assumptions of the present study is that every aspect of multimodal representation should be explicable within a single general cognitive model.

(i) *Attention/salience operations*

Focusing one's attention on the relevant aspects of our experience in the construal process, i.e. the activation of a specific profile within a semantic frame, is driven by the features that are salient in the given context. Let us recall the above example 4 of a cyclic gesture accompanying the verbal utterance *The wheel*

[28] In the following chapter, the cognitive theoretical framework will be discussed from a broader perspective.

[29] Again, this assumption has backgrounded Cognitive Linguistics since its earliest days (cf. Lakoff 1977).

is spinning. Such a multimodal utterance may result from various ways of profiling. Let us imagine that it was said by someone when describing how a water mill works. In the verbal part of the multimodal utterance, this profiling is manifested at the discursive level by focusing on the wheel's action before focusing on its appearance, for example. The gestural part of the multimodal utterance simultaneously profiles two aspects of the water wheel: its shape and the circular movement (rather than, for instance, its structure). Gesture can also embody the dynamic focusing of attention, highlighting "a construal of a static scene in dynamic terms" (*ibid*., p. 53) in the so-called fictive motion constructions (see below).

(ii) *Judgement/comparison operations*

This construal operation is based on contrasting several entities. Hand gestures allow for a simultaneous visualisation of separate conceptual entities in numerous ways: the simplest example could be the representation of two contrasted entities (objects, events, etc.) by both hands. Such a pattern, based on the OBJECT image schema (see below), seems to be very frequent in gestures accompanying constructions such as the adversative constructions (e.g. English [*not X but Y*]), or other types of constructions of comparison between two entities, e.g. [*either – or*] or [*On the one hand X – on the other hand Y*]. The last example is a special case of a directly iconic representation of a linguistic representation of the conceptual metaphor IDEAS ARE OBJECTS. Example 6 and Figure 8[30] show a speaker producing two gestures accompanying the spoken Czech utterance *bych mu to raději řekla* ***osobně****, než abych to psala někde* ***anonymně*** ('I would rather told him **personally**, rather than write it **anonymously**').

In this case, the contrast between two entities (two contrasting ways of telling bad news to somebody) is not visualized by the contrast between the two hands but by two bi-manual gestures with the same handshapes, produced first on the left-hand side of the speakers' gesture space and then on the right-hand side with a mirrored handshape configuration.

Recalling the above example of the gesture representing drawing in a sketchpad, we may see the two-handed gesture as an embodied representation of the FIGURE-GROUND construal. If, for instance, the speaker produces such gesture saying *pass me the pencil*, the fingers of one hand highlight the construed FIGURE and the open hand gesture of the other hand represents the GROUND (sketchpad), which is only implied.

Multimodal constructions similar to example 6 were addressed in the context of political speeches by Calbris (2008) and Miranda and Mendes (2015). Concern-

[30]The example comes from a corpus of recordings of narratives and dyadic interactions collected by the author in 2018 at Charles University in Prague.

Figure 8: *Contrasting between two conceptual entities*

(6) *Bych-∅* *mu* *to* *raději* *řek-l-a* *osobně* *než*
be.COND-1SG he.DAT it rather say-PST-1SG.F personally than
abych-∅ *to* *psa-l-a* *někde* *anonymně*
so.that.be.COND-1SG it write-PST-1SG.F somewhere anonymously

'I would rather told him personally, rather than write it anonymously'.

ing the systematicity in placing the items in the adversative relation on the right- or left-hand side of the speaker's gesture space, one might expect that the conceptual metaphor LEFT IS BAD, RIGHT IS GOOD (Lakoff and Johnson, 1980) would be reflected in the gesture. However, Casasanto and Jasmin (2010), who focused on the attribution of positive/negative value to the gesturally represented concepts in relation to handedness, found a tendency of gesturers to associate positive value with the gesture produced with their dominant hand (see also Section 3.2.1).

(iii) *Perspective/situatedness operations*

The construals of perspective involve situating subjects and their actions in a spatial and temporal context, with gesture naturally playing a fundamental role in these construal processes. I have already mentioned McNeill's distinction between the character and the observer viewpoint as well as the object- or action-centring strategies. Cognitive accounts of the use of gesture in viewpoint construal were provided by Sweetser (2012) or Parrill (2010). Viewpoint-strategies represent only a fraction of an immense spectrum of multimodal construals of perspective. Other types of perspective construals concern the *frames of reference*, temporal reference and other manifestations of spatio-temporal metaphor, which will be discussed in detail in the following section, dedicated deixis.

(iv) *Constitution/gestalt operations*

Besides profiling of certain qualities of conceptual entities (i), individuation and comparison between conceptual entities (ii), and the construal of the subject's relation towards conceptual entities (iii), gesture may be of course employed in the construal of the conceptual entities as such, i.e. the construa] of the "very structure of the entities in a scene" (Croft and Cruse, 2004, p. 63). According to Croft and Cruse, the concepts of *image schemata* (Johnson, 1987), *force dynamics* and *structural schematization* (Talmy, 2000) all correspond to constitution construals.

This notion of image schemata emerged in the early stages of CL, referring to "a recurring, dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience" (Johnson, 1987, p. xiv). Akin to image schemata, the notion of *force dynamics* was introduced by Talmy first in 1972 and later elaborated in his Cognitive Semantics (2000). Force dynamics is understood "the ways that objects are conceived to interrelate with respect to the exertion of force, resistance to force, the overcoming of such resistance, barriers to the exertion of force and the removal of such barriers, and so on" (Talmy, 2000, p. 219). Both image schemata, including the basic concepts such as CONTAINMENT, PART-WHOLE relation, BOUNDEDNESS, or SURFACE and force-dynamic schemata (e.g. CAUSATION, LETTING, REMOVAL, or ATTRACTION) stem directly from the human physical experience (see Section 3.2.1 for a discussion of embodied cognition) and are shaped by human sensory system, laterality of the body, the upright posture, and other somatic, kinetic and proprioceptive constraints (as such, these schemata are considered a universal basis for higher-level conceptual structures (conceptual metaphors, but also linguistic structures like prepositional systems or categories VERB or NOUN) that are subject to cross-cultural and cross-linguistic variance).

A considerable amount of work in the cognitively oriented studies of gesture has been carried out with focusing on gesture as the manifestation of image schemata or force dynamics. The idea of gestures as revealing such embodied schemata to direct observation was elaborated in numerous studies (e.g. Calbris, 2003, Cienki, 2005 or Mittelberg, 2018), under the assumption that

> "when gestural behavior is motivated by embodied image schemas and/or force gestalts, its core consists of inherently meaningful structures. […T]his kind of semantic essence […] should emerge from the corresponding gestures to a certain degree, even without considering the speech content" (Mittelberg, 2018, p. 3)

Calbris, for instance, described the recurrent gesture form (a variable but narrow set of phonological features) that mimics the action of cutting as a repre-

sentation of the image/force-dynamic schema of CUTTING – directly embodied and preconceptual, yet based on a complex network of sensory and proprioceptive experiences. Calbris shows how the base gestural schema is extended from clearly iconic instances (*cutting bread*) to metaphoric usages ranging over multiple domains (cf. *cutting the meeting short* vs. *cutting someone's funding*).

The conceptualization of how events unfold in time (the basis for the semantic domain of aspectuality) is based on the image schema of BOUNDARY, as well as force dynamic schemata related to characterise of motion within and across boundaries, underlie the conceptualization of how events unfold in time, which is a basis for the semantic domain of aspectuality. As the expression of event structure is one of the main foci of this study, a closer look will be taken at the multimodal representation of BOUNDARY schemata (Section 4.3.1).

Talmy's types of *structural schematization* (or *configurational structure*, 2000) encompass the different ways of construing the "topological, meronomic and geometrical structure of entities and their component parts" (Croft and Cruse, 2004, p. 63). In the example 5 (Figure 6), we have already seen how gesture may iconically highlight the internal structure of a conceptual entity. This example (a repeated cutting gesture along a horizontal axis accompanying the expression *rozfázovanou* ('divided up into phases')) is a case of a multimodal construal of what Talmy called the state of dividedness "a quantity's internal segmentation" (Talmy, 2000, p. 55), marked also by the distributive prefix *roz-*, and, at the same time, the pattern of distribution, "pattern of distribution of matter through space or of action through time" (*ibid.*), specified lexically by the stem *-fáz-* ('phase'). In the gestural representation, both structural configurations can be mapped onto the distinct phonological features: the representation of cutting corresponding to the state dividedness and the segmented horizontal movement to the pattern of distribution. However, it must be noted that it is not always possible to reliably associate gestural features with the individual aspects of a construal operation.

In particular, construals of the pattern of distribution as well as another structural configuration, the *state of boundedness*, distinguish specific categories within the grammatical-semantic domain of event structure and aspectuality. One of the key questions of this study concerns the role of gesture in the construal of these structural schemata and I will address them again in Section 4.3.2.

Construal is also central to a cognitive account of iconicity in the manual-visual modality proposed by Sherman Wilcox (2004). According to his model, called *cognitive iconicity*, "[i]conicity is not a relation between the objective properties of a situation and the objective properties of articulators. Rather, the iconic relation is between construals of real-world scenes and construals of form" (p. 123) within a single conceptual

space. Such a view corresponds to how gestural representation is approached in this study. There is no reason to assume that the principle of iconic mapping between semantic and phonological construals would be different in gesture and sign language signs. Hands as the primary articulators in the visual modality, allow for embodiment of conceptualizations via the modes of representation outlined above (Table 1) (see Wilcox and Morford, 2007) and the only real difference is, as already said above, the degree of conventionalization or lexicalization of these embodied representations. As we will see in Section 4.2.3, from the cognitive iconicity perspective, certain phenomena, such as representation of event structure across various sign languages, allow for direct comparison with the way event structure is represented in gesture.

Deixis

The third of the so-called "Iconic-Metaphoric-Deictic-Beat Quartet" (McNeill, 2006), are gestures with a primary function of direct or indirect reference. Deictic gestures may be used for referring to concrete entities physically present in the moment of utterance, displaced or abstract entities, including discourse topics. They are indeed often produced together with phoric expressions (e.g. demonstrative and personal pronouns), serving the same function in a co-expressive manner. Typically, deictic gestures are realized as *pointing*, but that is not the only way in which deixis is exhibited in gestures (see below).

In his book on the referential system of language, Leonard Talmy (2017) identifies a category of *targeting gestures* – gestures that "a speaker produces in association with a trigger specifically in order to provide a cue to a target" (*ibid*., p. 207) – and divides it further into *self-targeting* gestures and *outward targeting* gestures. Self-targeting gestures represent instances of meta-deixis – they refer to the very act of gesturing (the gesture constitutes the referent or in Talmy's terminology the target: the point in the discourse which is being referred to. A self-targeting gesture is thus a cue to the target and simultaneously the target itself. Outward targeting gestures refer to physically or temporally displaced referents. According to Talmy, outward targeting gestures are based on cognitive processes creating an imaginary link between the gesture and the target: so-called fictive chains. An inventory of fictive constructs – "topologically schematic abstractions" (p. 213) similar to image-schemata (see below) – through which the fictive chain is construed is supposed to be universal. However, the choices speakers of a particular language make when producing targeting gestures and hearers when they in turn construe fictive chains are subject to language-specific patterns. Fictive projection realized via fictive chains is in fact the manifestation of the same cognitive process that also exhibits itself in language in the case of fictive motion constructions like *The road goes along the coast* (Matlock, 2004).

The typology of deictic gestures proposed by Julius Hassemer and Leland McLeary (Hassemer, 2016; Hassemer and McCleary, 2018) is complementary to

Talmy's treatise of gestural targeting. Their proposal is based on seven spatial cognitive operations (*articular profiling*, *shape profiling*, *extension*, *intersection*, *trace leaving*, *limiting*, and *proximity*) that can be variously combined – 27 described combinations lead to specific types of pointing gestures, falling into a macro-category according to the profiling of the hand[31] – either as a *vector* or a *surface*. These two macro-categories have a partial affinity to Talmy's self- and outward targeting gestures: vector-profiling gestures are gestures of pointing to a distant target (where the vector is the fictive chain), whereas surface-profiling gestures point to the articulators like self-targeting gestures. However, unlike self-targeting gestures, Hassemer's and Leary's types like SURFACE-EDGE MARKING or SKETCHING WITH SURFACE do not represent only the act of gesturing itself; they also refer to something else. This leads us to a very important point that Hassemer and McLeary make: pointing gestures can serve both deictic as well as *iconic* functions.[32]

Gestures like the just mentioned are characteristic examples of this semiotic conflation (in fact, they would likely only be labelled as iconic by some researchers). Vector-profiling gestures can often be considered iconic as well. In this case, the articulator (e. g. index finger or open palm) is profiled as a vector, not a surface, and as such does not depict anything – however, a fictive extension of the articulator leading to the referent (an imaginary ray emanating from the articulator), becomes an iconic depiction of the fictive chain.

In a similar manner to beat gestures, deictic gestures are also realized by different body parts. Apart from fingers, deictic gestures may be produced by rotating one's head, directing the eye gaze, pointing with a foot, pointing with the lips, or positioning of the entire body – the list might go on as the deictic function can be construed ad hoc with any tools at hand.[33]

Pointing is sometimes considered a rudimentary kind of gesture – from both ontogenetic and phylogenetic perspectives. It seems that communicative pointing in prelinguistic children is a culturally universal phenomenon (Liszkowski et al., 2012). First fully-fledged pointing gestures appear on average at 11 months of age (Butterworth, 2003), i.e. slightly before the first isolated words are produced (12 months). However, some form of pointing behaviour can be observed in children as young as 3 months (*ibid.*).

Pointing gestures are also where human and non-human communication ex-

[31] Although not explicitly limiting their discussion to hands, Hassemer and McLeary almost exclusively deal with manual gestures.

[32] Hassemer and McLeary quote Charles Goodwin's remark in a similar vein: "In most typologies of gesture […], iconic gestures and deictic (pointing) gestures are treated as separate kinds of gesture. This does not seem to be correct. Pointing gestures can trace the shape of what is being pointed at, and thus superimpose an iconic display on a deictic point within the performance of a single gesture" (Goodwin, 2003, p. 229).

[33] Consider, e.g., possible pointing by throwing objects or by actions such as spitting pits – which just highlight the fictive chain in a more direct way.

hibits the most similarities. The primate pointing has long been at the forefront of animal communication research and although apes indeed point in a way resembling what humans do, it was assumed that, unlike human children, apes do not develop declarative pointing (i.e. for the purpose of establishing joint attention) – they only use pointing gestures in an imperative manner (Tomasello, 2005). More recent evidence suggests that even this higher social function of gesture is, in fact, not exclusively human and can be found in ape gesturing too (Halina et al., 2018).

There is a similarity of form and function of pointing gestures to pronominal signs (Cormier et al., 2013) or so-called agreement (or indicating) verbs (Liddell, 2000) in many sign languages. When approached as potential multimodal constructions (see 3.2.3 for discussion of Multimodal Construction Grammar), agreement verbs and pointing gestures share the core features but differ in the degree of grammaticalization (Schembri et al., 2018). However, in some spoken languages, pointing gesture patterns also exhibit characteristics of grammaticalization (or constructionalization), as illustrated below.

Deictic gestures have also been paid attention by linguists in terms of cross-linguistic variation. Wilkins (2003) argued against the claims of the universality of pointing with an index finger. Whereas it may be the primary means of gestural deixis in Western cultures, in other parts of the world, pointing with one's lips may be used to the same extent, or even completely at the expense of manual pointing.[34] Regarding manual pointing, Wilkins also presents evidence for a diversity of different handshape configurations that can have conventionalized functions. Speakers of Arrernte, a language spoken in the Northern Territory of Australia, use a number of different handshapes for specific kinds of pointing, directions and spatial depictions. For instance, pointing with a handshape with extended index and little finger (similar to the "horns" – a common emblem in many Western cultures) conveys the meaning of a "global orientation of a place that is being moved to, independent of the orientation of the subpaths used to get there" (*ibid*., p. 185). Cooperrider, Slotta and Núñez (2018) provided further evidence for cross-cultural differences in the preferred form of pointing. In their experimental study, the authors compared groups of American English speakers and speakers of Yupno (a Papuan language) and observed that under the same conditions, both groups produced a similar amount of pointing, however the Yupno showed a strong preference for non-manual pointing (nose and head gestures), whereas the American group used almost exclusively manual pointing.

It is not only the form of pointing, that varies across cultures, but it is sometimes also the case that the same form may serve different functions. Floyd (2016) describes how speakers of Nheengatú, a language spoken by small communities in the Amazon, use pointing with an extended index finger for referring to specific times of day. Together with temporal expressions, Nheengatú speakers point to the corresponding

[34] E.g. in certain speech communities in the Bird's Tail Peninsula of Papua New Guinea.

position of the sun in the sky.[35] This way they can also specify the temporal reference when the verbal means are too vague and provide a coarser estimation. The Nheengatú are able to locate the appropriate position of the sun quite precisely regardless of the current time of day, visibility etc. Such ability can be explained as an example of relativistic effects of how spatial-temporal relations encoded in specific languages may influence non-linguistic spatial abilities in speakers of those languages (Levinson, 1996b).

Linguistic relativity effects in the domain of space are materialized in referential gestures also in case of *frames of reference* – linguistically encoded coordinate systems grounding the use of location expressions (Levinson, 1996a). Levinson (2003) discusses the use of gestures with spatial language in Tzeltal and Guugu Yimidhirr, two languages known for the preference of (variants of) an absolute frame of reference (*ibid.*).

Three frames of reference for describing spatial situations are distinguished: *absolute* (location based on fixed bearings, typically cardinal directions, e. g. *the porch is north of the house*), *relative* (location relative to the speaker, e. g. *the porch is to the left of the house*), *intrinsic* (location based on the inherent features of the ground, e. g. *the porch is in front of the house*). Languages differ in which frame of reference is preferred for describing similar configurations of FIGURES, GROUND and OBSERVER.

In Tzeltal (Mexico), situations which would be described using relative and intrinsic frames of reference, e.g. in Indo-European languages[36] are predominantly coded in an absolute frame of reference. In this case, the coordination system is not cardinal but landmark-based, reflecting the site-specific topology: the lexically encoded coordinates being *ajk'ol* ('downhill') and *alan* ('uphill'). Interestingly, there are no dedicated terms for LEFT and RIGHT, while the Tzeltal absolute system operates with laterally non-specific *jejch* ('across').

Tzeltal speakers were reported to use the absolute frame of reference systematically in their spatial descriptions even when displaced, including precise use of pointing in accordance with the coordination system. This ability to maintain complex spatial reference even without visible landmarks, as if being endowed with some sort of "inner compass" is often presented as one of the soundest examples of linguistic relativity (*ibid.*).

Speakers of Guugu Yimidhirr (Cape York Peninsula, Australia), another language with an absolute frame of reference (in this case based on cardinal directions) exhibit the same tendencies. As Levinson points out, "absolute 'semantics' seems to pervade the production and interpretation of all the gestures in such a system – not just pointings, but depictions or 'iconic' gestures as well" (Levinson, 2003, p. 248).

[35] Celestial pointing is used for temporal reference in sign languages too, e.g. in Kata Kolok, a village sign language in Bali (de Vos, 2015).

[36] Or "Standard Average European" (Whorf, 1941).

Beat gestures as iconics?

In their initial definition, McNeill and Levy described beats as "small rapidly made gestures with indefinite form that do not depict any aspect of the verbally described situation" (1982, p. 285). However, as has been pointed out above, that does not mean that beats do not have any relation to the structure of the accompanied utterance: it is only that they do not seem to be associated with semantics at lexical level. McNeill later expanded on beats:

> "They are mere flicks of the hand(s) up and down or back and forth that seem to 'beat' time along with the rhythm of speech. However, they have meanings that can be complex, signaling the temporal locus in speech of something the speaker feels is important with respect to the larger discourse" (McNeill, 2005, p. 40).

The function of beat gestures in the marking of discourse segments is directly linked to the relation between the production of beats and prosodic properties of the accompanied speech. This is one of the aspects of gesture that had attracted the attention of linguists long before the birth of gesture studies as a discipline (Bolinger, 1983; Pike, 1967, *inter alia*). Although the assumption that gesture and intonation are coordinated had been acknowledged before, it was Adam Kendon who, in his pioneering analyses of the use of gesture in interaction (1972; 1980), first demonstrated that (i) co-speech gestures are organized into phrasal units in a similar manner to prosodic units and that (ii) gestural and prosodic units align. According to Kendon, a gesture should not be considered to be an insignificant movement in the case of beats or a "parallel" symbolic system in the case of emblems since gesture is, alongside speech, an integral component of an utterance (as discussed in Section 3.1.1).

From the production perspective, this dual accentuation has been explored in a number of studies. In one of the first quantitative studies of the relation between gesture and prosody in English spontaneous production, McClave (1998) found a general tendency for gesture strokes to co-occur with "fundamental frequency in the intonation group" (*ibid*., p. 87). Ferré (2010) focused on the alignment of corresponding speech units with iconic gestures in French spontaneous production. Having obtained data from the CID corpus (Bertrand et al., 2009), Ferré compared temporal alignment of iconic gestures and affiliated words as well as coordination between entire gesture phrases and affiliated intonational phrases. At both levels, she found a significant majority of gestural units appearing before the onset of the corresponding phonological unit. Moreover, there was no case of a gestural unit ending before the end of the corresponding phonological unit. This temporal shift had been observed before and was most notably addressed by Loehr (2004). Along the same line as Ferré's findings from French, Loehr's results revealed this "duality of patterning" within the gesture-speech interface in English: at the level of gestural phase/basic intonational unit as well as at the level of gesture phrase and intonational phrase. Since Loehr applied more fine-grained coding to his data, he identified the specific loci of the alignment: the gestural

apex with pitch accent and the gestural phrase with intermediate phrase. The intermediate phrase thus might be, according to Loehr, the phonological analogue of the gesture phrase, both "correspond[ing] to the size of the cognitive package which can be expressed through a single surge of bodily and vocal action" (Loehr, 2012, p. 85).

The gesture-speech orchestration is not hierarchical: effects of bidirectional influence between the two parts of the interface were reported (Pouw et al., 2020b, or the studies of cross-modal effects in gesture-speech comprehension mentioned in Section 3.2.2). At the most general neurological level, this interface is likely to be related to the topological adjacency of the speech production and manual activity centres in the brain (addressed in Section 3.2.1). Another piece in the puzzle might be *respiration* as breathing is coordinated with both the movements of the upper limbs and suprasegmental qualities of speech (Pouw et al., 2020a).

The discursive function of (beat) gestures was first addressed by Kendon in his ethnographic analysis of Neapolitan gestures (1972). He observed and described recurrent gesture forms that mark topical information. McNeill (2005, McNeill et al., 2015 follows Kendon's integrative view of gesture and sets it into the framework of cognitive linguistics. Thus, according to McNeill, gesture and speech are not only two parts of an utterance but also two means of expressing the conceptual content of the speakers' minds. According to McNeill, gesture serves as a means of creating and maintaining the cohesion of discourse. Different segments (e.g. different topics) of discourse may be distinguished using different types of gestural forms. McNeill calls these discourse-gestural units *catchments*: "thread[s] of consistent visuospatial action imagery running through the discourse and [providing] a gesture-based window into discourse cohesion" (McNeill et al., 2015, p. 267). Within a catchment, smaller units – basic meaningful gesture-discourse mappings (*growth points* – discussed in detail in Section 3.2.2) – can be recognized and they differ with respect to their communicative dynamism (McNeill adopts the term from Jan Firbas' (e.g. 1992) theory of information structure (IS)) i.e. in the extent to which they drive the discourse forward.

In the data analysed by McNeill and Levy (1982), the distribution of gesture-discourse units correlated with the distribution of prosodic units. Catchments were found to be prosodically bounded – in accordance with Kendon's findings mentioned above. On the boundaries of those gesture-prosodic segments lie the gestures introducing a new catchment, in other words, units with high communicative dynamism. Having illustrated how speakers embody the information structure of the discourse into gestures they use, McNeill et al. came to the conclusion that

> "the organization of discourse is inseparable from gesture and prosody: the three components are different sides of a single mental-communicative process. A purely text-based approach, as in the narratology tradition, is blind to two-thirds of this discourse structure" (McNeill et al., 2015, p. 273)

Yoshioka (2008) compared the use of gesture in topic-marking in Japanese and Dutch.

In her analysis of gesture use in elicited narratives, she found that gestures are more frequently used to mark the information status of referents (given vs. new) in Japanese than in Dutch. This finding might reflect the fact that Dutch employs more grammatical features for marking IS than Japanese (such as articles, greater variety of pronouns, etc.).

Using a corpus of recordings of spontaneous English production occurring at public town hall meetings, Jannedy and Mendoza-Denton (2005) focused on how pitch accents and gestures align in order to structure the information in discourse. For capturing the intonation information, Jannedy and Mendoza-Denton used the ToBI annotation scheme (*Tones and Break Indices*, Silverman et al., 1992). Results of the microanalysis revealed that every gesture apex in the analysed extract co-occurred with a pitch accent – marking informationally salient content of the speaker's production. Since the authors' approach was based on the conversation analysis framework, the scope of their analysis is limited (130-second extract containing the production of two speakers).

A study by Ebert et al. (2011) represents a corpus-based analysis of a larger scale. Investigating gesture-pitch accent alignment with sentence focus, the authors analysed data obtained from the SaGa corpus (Lücking et al., 2010), which contains 280 minutes of video-recorded dialogues between 25 dyads of speakers elicited using a direction-giving task, including annotation of gestures based on several aspects of gesture forms and transcription of the speech. The authors selected one interaction (20 minutes, two speakers) and provided the data with additional annotation of prosodic features based on the ToBI framework and annotation of IS, namely contrastive and new-information focus, according to IS-annotation guidelines by Dipper et al. (2007). Results of the analysis of temporal co-ordination of more than 250 instances of focus-pitch accent-gesture alignments revealed that the onset of a gesture phrase systematically precedes the onset of the corresponding focal unit. Evidence for gestural marking of various IS units was reported in Turkish (Turk, 2020), American English (Im and Baumann, 2020) or Czech (Lehečková et al., 2019).

From the Peircean perspective, a prototypical way of gestural marking of prosodic/discourse structure in the languages mentioned above constitutes an iconic mapping of a diagrammatic kind between the segments of gesticulation marked by stroke apices and intonational phrases marked by F0 peaks. But could we consider gestural iconicity also in the case of beat gestures affiliated to constructions at the level of lexical unit or a syntactic phrase?

Beats are sometimes treated as a sort of "garbage bin category" where all other gestures that cannot be easily labelled as iconic, deictic or metaphoric, are tossed. Framing beats as "mere flicks of hands" may be motivated by the apparent lack of complex phonology that could be mapped on semantics at the lexical level.

Recently, a number of studies emerged that challenge the simplistic view of beat gestures (Ruth-Hirrel and Wilcox, 2018; Prieto et al., 2018), shifting attention to the

complexity of pragmatic functions of beats beyond discourse-marking. However, none of these studies truly tackled the semantic potential of beat gestures, i.e. reference to the qualities of conceptual structures represented linguistically at the level of lexical or idiomatic constructions.

But even the apparently purely rhythmic beats (i.e. the small gestures without articulated handshapes, swinging rhythmically up and down or back and forth, allegedly "communicat[ing] nothing specific beyond emphasis" (Abner et al., 2015, p. 438)), may exhibit a significant variation in terms of the phonological feature manner of movement. Specifically, simple beats (however rarely they may actually occur in the simple form) may vary in the parameters such acceleration or presence of accentuated ending (Bressem, 2013). One of the arguments of this study is that this variation in gesture production (and modulation of the said parameters in a perception experiment) is systematically associated with the linguistic encoding of event structure. As the results of both a corpus (Chapter 5) and experimental study (Chapter 6) will illustrate, variation in these parameters can be mapped on specific semantic features of the accompanied speech.

3. Gesture and cognition

> *"Larger social realities are built up from thousands and thousands of small-scale interactions, so a social scientist's need to understand these tiny moments is a bit like a physicist's need to understand subatomic particles"* (Dingemanse and Floyd, 2014, p. 447).

In the mid-1920s, German physicist Werner Heisenberg published an article (1927) in which he famously stated that when measuring the properties of subatomic particles, it is impossible to account for all variables (e.g. velocity, position or energy) at once. Given their wave-like nature, subatomic particles – basic building units of the universe – cannot be captured by the observer in terms of all physical parameters. This finding was later dubbed the *uncertainty principle* and resulted in one of the great scientific schisms of the 20th century. It is rooted in a paradox: the same laws of physics do not appear to hold for both the micro- and macroscopic level. Another aspect[37] of this principle concerns the observation itself. Even at the microscopic level, the very act of observation affects the way elementary particles behave (e.g. measurement disturbs the system of particles by introducing interacting photons), thus making it effectively impossible to observe them in their unobserved, or "objective" state. What remains available to the researcher is only their interpretation. This idea is known as the *Copenhagen interpretation of quantum mechanics*. With its paradoxical nature, this interpretation has always caused conflicts amongst theoretical physicists, with the most notable one being, arguably, the dispute between Niels Bohr and Albert Einstein in the 1930s. The clash between quantum mechanics and relativity theories that hold for the physics of micro- and macroscopic worlds respectively, but are fundamentally incompatible with each other, became a battleground not only of theoretical physics but philosophy of science as such. However, even after the battles of old passed, the schism is still present in contemporary physics: the search for a unified theory reconciling both worlds is not over yet (Kragh, 2002).

William Labov who acknowledged the observer's paradox in linguistics, pointing out that, on the one hand, the desideratum is to capture "that every-day speech in which the citizen scolds his children, jokes with his friends; and orders a slice of apple pie" (Labov, 1964, p. 167). On the other, the presence of a researcher with their recording instruments inevitably disrupts the equilibrium of spontaneous speech production, "It is the familiar problem of whether the light is on or off when the refrigerator door is closed" (ibid.).

[37] Although Heisenberg does not conflate the uncertainty principle with the observer effect in his 1927 paper, as is sometimes incorrectly assumed. However, both concepts are related and mentioning them together is legitimate in the light of Heisenberg's own later remarks (1958).

Following this metaphorical frame, one might argue that, hyperbole notwithstanding, the quantum-mechanics – relativity schism could be likened to the relation between approaches to the study of how language works in the interaction of individuals on the one hand, and in the framework of general theories of human cognition on the other. Labov's call for collecting linguistic material in such a way that causes the least possible detriment to spontaneity and casualness of production followed by a transformation into a objectified and generalized set of *data* still resonates in the current methodological discussion with particular relevance for the study of gesture-speech integration. The following sections will be framed by theoretical and methodological tensions stemming from this call.

If we focus on gesture alone, so far we may have gotten the impression that although it is often challenging to describe its functions and semantics, it may be viewed as a physical unit with clear boundaries, analysable on its own. This is how gestures are typically treated in cognitivist studies. Partly, such view is quite fitting for the co-speech gestures occurring in the type of material typical for this area: narrations in highly controlled settings, elicited to produce clear-cut instances of (mostly) iconic gestures.

Nevertheless, when we turn to gesture in spontaneous everyday interaction, we are often faced with a quite different phenomenon. Not only is it difficult to attribute a function and meaning (or a limited set thereof) to a gesture, it also frequently becomes apparent that it cannot be established precisely where the gesture (*any* given gesture, even the clear-cut instances mentioned above) begins and where it ends.[38] Gesture can be viewed as "particle-like", and at the same time it also is fundamentally "wave-like" – thus always fleeting and inaccessible to a full and definitive account.[39]

However, the heart of the matter may lie somewhere else than in methodology alone. As J.P. de Ruiter and Saul Albert pointed out, the root of the schism is more profound, and just like in the above case of physics, it pertains to the fundamental assumptions about how and from what kind of evidence scientific knowledge originates (de Ruiter and Albert, 2017; I will discuss this in Section 3.3).

One of the paradoxes underlying the Copenhagen interpretation of quantum mechanics is the so-called dual nature of matter: should the subatomic domain be modelled upon entities that are particle-like or wave-like? Niels Bohr (1928) concluded that both models are correct, given appropriate circumstances. Regardless of the adequacy of this particular theory (which is more of an issue for the historiography

[38] On the issue of recognition of sign language signs see Jantunen (2015).

[39] Even though when we compare it to the problem of the reconciliation of the two worlds of theoretical physics, the problem of the two approaches to communication does not seem to pose such a challenge for the philosophy of science. The fundamental disagreement seems to be purely methodological and does not concern the nature of the phenomenon in question – after all, both sides probably agree upon the basic assumptions about language. That is why there is no real "discursive clash" but rather two disjointed paths, peacefully but mutually neglectfully coexisting, both equipped with a different set of descriptive and analytical tools.

of science nowadays), the logic behind the complementary principle may provide the key to overcoming the interactionist – cognitivist schism. Rather than conceptualizing them in a dichotomous manner as mutually exclusive methodological alternatives, we could view the two approaches as related to different domains of scientific scrutiny: each follows different principles of data acquisition, but that does not imply that the data are incompatible. On the contrary, there could be a complementary relation between the two types of evidence if appropriate attention is paid to what has been called the "context of discovery" (3.3).

The goal of this chapter is to introduce a theoretical framework that supports the interpretation of the results of the present empirical study – a framework that assumes the complementarity of the interactionist and cognitivist paradigms.

In the subsequent sections, I aim to postulate a methodological dichotomy within mainstream gesture studies. This dichotomy reflects a general theoretical-methodological divide between the study of human communication based on close observation of human interaction in various settings on the one hand, and cognitive approaches to communication based primarily on experimental methods on the other.

For simplicity's sake, I present the relation between two major methodologies in gesture studies as a kind of a "clash of paradigms". But in fact, there has never been an explicit dispute[40] among gesture researchers in this regard. Rather than competing, the two approaches have coexisted peacefully, manifesting their mutual distance by not cooperating.

I start with the *interactionist* approaches (3.1), focusing on major concepts introduced by Adam Kendon (3.1.1) and Jürgen Streeck (3.1.2).

A closer look will be taken at the *cognitivist* approaches (3.2). First, I will provide a summary of the key lines of evidence related to gesture's role in human cognition (3.2.1). Then I will review a model of language processing integrating gesture (3.2.2) and subsequently, I will introduce one of the current cognitivist frameworks for analysing multimodal expressions developed from Construction Grammar (3.2.3).

This chapter is concluded by a proposal for an integrative approach to the study of gesture, bringing together "the best of both worlds" – highlighting ways in which interactionist and cognitivist views are complementary (3.3).

3.1 Ecology of gesture: an interactionist approach

From the perspective of today's gesture studies, it is the anthropologist and Franz Boas'[41] student David Efron who is often pointed to as the first one to take up the

[40] Unlike the case of theoretical physics mentioned at the beginning of this chapter.

[41] It is worth noting that Franz Boas in fact expressed his intention to record gestural behaviour of North American native peoples and planned to use a camera to capture it. Eventually, this pioneering project was completed only partially due to Boas' death in 1942 (Ruby, 1980) – and it had taken more than 30 years until filming techniques gradually started to be employed in language documentation.

matter of co-speech gesture as a subject of scientific scrutiny with its own descriptive apparatus (Kendon, 2004). Efron's work, although not continued by immediate followers, was a vanguard of the study of gesture in interaction. In his ethnographic description of gestural behaviour of two groups of immigrants with different cultural backgrounds, Efron (1941) provided the first systematic account of the linguistic nature of co-speech gestures, suggesting that they are not mere meaningless movements but rather serve to emphasize the content expressed verbally. Efron's work, apart from its ethnographic novelty,[42] is remarkable as it foreshadows the future interactionist research. It did so not only by close observation of naturally occurring gesticulation, but it also pioneered the ethnographic technique by combing field notes and cinematography. Efron's observational data were even subjected to a rudimentary quantitative analysis, limited in its scope but, as noted by Adam Kendon, "still large by comparison with the numbers of individuals studied in most observational studies today" (2004, p. 331). Another aspect in which Efron's work prefigured the modern ethnomethodology is the focus on the inter-party character of gesture production and its situatedness in an actual physical setting (cf. what Charles Goodwin later called "contextual configuration" (Goodwin, 2000).

Another inspiration for the interactionist stream in gesture studies is represented by the phenomenological works of Maurice Merleau-Ponty. The core of his position on language is formulated in *Phenomenology of perception* (*Phénoménologie de la perception*, 1945)[43] which is usually also his only work that is referred to with respect to language or gesture. Gesture plays a prominent role in Merleau-Ponty's approach to language: "The central theme of [Merleau-Ponty's reflections] when it comes to speech is the reflection of gesture, for it is an extreme phenomenon situated between the body and linguistic expression, being at the same time a semiotic unit and a bodily action." (Fulka, 2017, p. 44) The gist of Merleau-Ponty's conception of language may be labelled as a *language-as-gesture thesis*. In his view, linguistic expression is nothing but a kind of bodily expression and as such is gesture-like. As for gesture itself, it has to be made clear that for Merlau-Ponty it means virtually any expressive act of communication, be it verbal language, manual gesture or other bodily movements including ritualized actions, e.g. dance. Such an extended sense of gesture is not entirely alien to linguistics: in articulatory phonology, the term *phonetic gesture* refers to a setting of articulatory organs that represents a basic structural unit (Browman and Goldstein, 1992).

As I have already noted, the critical moment in Merleau-Ponty's serving as an inspiration for gesture studies relies on his "anti-mentalist" position. However, his stance cannot be simply labelled as *behaviourist*. Whereas for strict behaviourism (Skinner, 1957) mental states are a taboo, Merleau-Ponty attempts to address them by means of

[42] An early focus on how gesture is shaped by culture and the pioneering observations concerning speakers' accommodation of gesture production in new environments.

[43] In this text, I cite the English translation (Merleau-Ponty, 1962).

their "materialization" through *embodiment* (*incarnation*). Therefore, the mental "content" can be indeed dealt with, though only figuratively.

After a considerable hiatus, Efron's pioneering work was revived in the 1960s. This new interest in gesture emerged as a part of the turn to the pragmatic and interactional aspects of communication (1962) and Searle (1969), i. a., on the part of philosophers, Hymes (1967), Grice (1975) or Leech (1983) on the part of linguists) that echoed in the emergence of Conversation Analysis (CA, Schegloff, 1968). As CA focuses on the communicative patterns of speakers in everyday interaction (capturing verbal as well as nonverbal linguistic behaviour with its notation system), it naturally lends itself to the inclusion of gesture into the analysis too.

The very beginning of gesture studies as a discipline that would eventually become embedded within linguistics thus roots from the interactionist paradigm that does not draw attention to language as a system as such, but instead aims to describe the pragmatic aspects of language use in specific situations and/or for specific purposes. As such, CA does not provide linguists with an *explanation* of language (verbal as well as nonverbal) use, but restricts itself to a thorough ethnographic *description* of context-grounded excerpts of language practices, avoiding generalization beyond the situation-specific frame.

CA equipped gesture studies with a key methodological tool that has been - with modifications - widely used, also beyond CA, until today: the transcription system developed (primarily) by Gail Jefferson (1984/2004). The Jeffersonian notation introduced a systemized and convenient way of capturing a multiplicity of properties of speech (at the suprasegmental level) and – crucially - the sequential and simultaneous nature of talk in interaction, allowing for monitoring overlapping speech, imperfections and repairs. In its original form, the Jeffersonian system did not allow for a fine-grained description of gestural behaviour apart from gestures unaccompanied by speech. But due to its extendability by means of layers (additional lines in the transcript), cco-speech gestures can be captured at a more sophisticated level (mainly when applied in a modern annotation software based on layered time-aligned notation, such as ELAN (Wittenburg et al., 2006). For a thorough discussion of gesture notation in the CA framework, see Bohle (2014).

Setting the detailed discussion of the CA-oriented gesture studies aside (the orthodox form thereof can be found in the works of scholars such as Charles Goodwin or Lorenza Mondada) I will further address two concepts that stem from the interactionist tradition and that have had an impact on gesture studies in general, having been embraced beyond CA.

3.1.1 Gesture as utterance

The early work of Adam Kendon (1972), initiated in the interactionist framework, provided a significant insight into the general aspects of gestural behaviour and may be

considered a pivotal point in the establishment of modern gesture studies (McNeill, 2005, p. 13). Kendon's contribution to the field has been immense, as he shaped much of the agenda of the study of gesture during its initial stages. Kendon's core idea is based on understanding gesture as a realization of what he calls *utterance*, i.e. "any action or ensemble of actions that may be employed to provide expression to something that is deemed by participants to be something that the actor meant to express, that was expressed wilfully" (Kendon, 2004, p. 8). Such a view allowed for gesture (or generally any *visible bodily action*, including posture, eye gaze, etc.) and speech to be analysed not as a hierarchical system, but as two equal phenomena. Both are forms of utterances, and although they contribute to the overall message in a specific manner (cf. McNeill's concept of the dual nature of imagery in speech and gesture discussed in 3.2.2), gestures are analysable in terms of units that engage with speech (and other utterance-generating actions) at various levels. Thus, the gestural units addressed in Chapter 2 were brought forth: gestural phases and phrases – the very fundamentals of the gesture studies analytical toolbox.

In his cognitive theory of gesture, David McNeill was substantially influenced by Kendon's insights. While adopting the core ideas, however, McNeill focused in particular on the nature of imagery and conceptualization that underlies the *representational* function of gestures, one of the main functions recognized by Kendon, who also (and above all) paid attention to pragmatic and discursive roles that gestures play in interaction. Another integral aspect of Kendon's approach sometimes becomes lost in the transition between Kendon and his non-interactional epigones; namely, he consistently takes into account the dialogical nature of utterance-making out of focus, always considering the multiplicity of parties involved.

The idea of utterance as a multimodal assemblage of expressive means has been widely adopted in gesture studies, in both interactionist as well as cognitivist streams. Nick Enfield, a researcher crossing over the two approaches, introduced an elaboration called *composite utterance* (2009). In his view, meaning is a layering of conventional as well as non-conventional signs that should not be understood in the linguistic sense – they originate from the actor's intentions in social interaction in general. The layers – signs – are in a dynamic relationship and they can only be analysed in a top-down manner: the meaning of a particular sign (e.g. the gestural or spoken component of the utterance) can only be interpreted with respect to the entire composite utterance. Enfield introduces the notion of *enchrony*, which refers to the temporal perspective spreading over "data from neighbouring moments, adjacent units of behaviour in locally coherent communicative sequences (typically, conversations)" (Enfield, 2009, p. 10). This perspective entails a basic analytic unit for composite utterances, a *move* of social action, corresponding to the conversational turn as it is understood in CA.[44] Enfield's account is a modality-free account – in terms of utterance one may describe

[44]NB that Enfield's approach is in many respects compatible with the Interactional Construction Grammar – see 3.2.3.

both spoken and sign language production using a semiotic approach independent of modality-specific notions (this is particularly important due to the implicit bias to resort to spoken language when talking about meaning).

3.1.2 Ecologies of gesture

Following in Kendon's footsteps, Jürgen Streeck is another key figure in the interactionist wing of gesture studies. The scope of Streeck's investigation of gesture is not limited to co-speech gestures. Rather than deriving his conception of gesture from general or language-related cognitive functions, Streeck approaches gesture as a *tool* for communicating meaning *sui generis*:

> "I [conceive] of gesture neither as a sign-system nor as a part of language (or a 'body language' onto itself) nor as expressive behavior (behavior that reveals what goes on in the person's mind or psyche), but as a mode of communicative praxis and craft, comprising skills, methods, and technique" (Streeck, 2009, p. 203).

Streeck distinguishes particular aspects of how gesture may be involved in communicative practice. He recognizes six of these aspects which he calls *ecologies of gesture*, "a distinct pattern of alignment between human actors, their gestures, and the world" (ibid., p. 7): (1) making sense of the world at hand, (2) disclosing the world within sight, (3) depiction, (4) thinking by hand: gesture as conceptual action, (5) displaying communicative action, and (6) ordering and mediating transactions (ibid, *passim*). Crucially, and in contrast to other functional classifications, the *ecological* (or "praxeological" – building upon Marx's and Vygotsky's understanding of the term *praxis*) perspective views gesture not as objectified communicative behaviours of an individual (an approach following, as Streeck points out (2009, p. 14) the Wundtian tradition – see 3.2), but as intersubjective phenomena emerging in interactions (a view inspired by G. H. Mead): "Gesture is not in the first place a means of expression; it is a component of social acts" (ibid., p. 15). Embedded in social acts (or practices), gesture does not allow for laboratory dissection detached from its intersubjective and environmental habitat; instead it must be observed "within real-world practice communities, from grocery stores to butcher shops and the workshops of tailors and blacksmiths" (ibid., p. 30).

Such a dynamic view of gesture as a process of situating a subject in the context of interaction, being inseparably linked to the experience with the real world and other subjects through bodily action, is explicitly linked to the tradition of phenomenology. It is so not only with regard to Mead's sociology and Merleau-Ponty's idea of gesture as a bodily expression (which will be discussed below), but it is also reflected by the term *ecology*, echoing another key concept of continental philosophy, namely that of *Umwelt* (Uexküll, 1992).

It adheres to the phenomenological approach to gesture as action and methodological principles of CA: the inseparability of the described action from the actual setting in which it is situated and the dissolution of the dichotomy between an individual instance and evidence-based generalization. "[T]he individual more or less consists of an accumulation of sociocultural 'stuff' (concepts, words, knowledge, beliefs, etc.) and sociocultural methodologies […] that are 'enacted' or 'used' in moments of social life" (Streeck, 2009, p. xx).

The interactionist research of gesture, while to a certain degree obscured by particular studies following the dogmata of classical CA that somewhat impose a kind of hermeneutical ring around its object of scrutiny due to the inherent scepsis towards induction (de Ruiter and Albert, 2017). Nevertheless, it yielded indisputable contributions, which proved to be necessary for the development of of gesture studies. After all, it was the interactionist approach that gave birth to the very discipline as we know it today. Also, the fundamental notions concerning gesture analytical units or functional classification were formulated based on observations of gesture "in the wild", not in the psycholinguistic laboratories, which in turn produced much valuable evidence concerning a much narrower spectrum of phenomena in which gesture is manifested.

3.2 Cognitivist approach to gesture

The other major approach to linguistic study of gesture emerged alongside the cognitive linguistics (CL) movement (Lakoff and Johnson, 1980; Lakoff, 1987) in the early 1980s in the United States. As such, the emergence of the CL-based approach to gesture was one of the results of the so-called Second Cognitive revolution.[45] Moreover, gesture's crucial role in our understanding of how cognition functions has been highlighted since the first formulations of the CL foundations appeared (McNeill, 1979; McNeill and Levy, 1982). Looking further back, a major precursor of the cognitive approaches to gesture may be traced, similarly to David Efron and his early study of gesture in interaction. Preceding Efron by several decades, the work of Wilhelm Wundt (1832–1920), a German psychologist traditionally considered the founding father of experimental psychology, is of relevance with respect to the cognitivist stream within contemporary gesture studies. Somewhat surprisingly, relatively little attention has been paid to Wundt in this regard.

The monumental *Völkerpsychologie* (1900b) includes a bulk of what Wundt wrote on language, which was the main topic of the first and second volume of the ten-volume series. Before including the two volumes into the *Völkerpsychologie* project, Wundt had published them as *Die Sprache* (1900a). Concerning gestures, the first and in particular the second chapter of the first part of *Die Sprache* are the most relevant parts, with the first one dedicated to "expressive movements" (*Ausdrucksbewegungen*)

[45] Or, more appropriately, "Second Generation of Cognitive Science" – see discussion below.

and the second one to gestural communication (*Gebärdensprache*).

In Wundt's view, gesture is a type of expressive movement, i.e. behaviour in which the mental states (*Gefühle*, i.e. emotions, which are of particular interest in the chapter) of a person are manifested. The way emotions – or more precisely affects (*Affekte*), i.e. changes of emotional states – unfold is reflected in various types of gestural behaviour (*Ausdrucksbewegungen*). The expressive movements do not only entail gestures but subsume all kinds of expression, including spoken language itself.[46] To this day, Wundt's classification of what he called "affective gestures", i.e. gestures that are motivated by affects, remains imprinted in the standard functional classification (2.2). The classification is schematized in Figure 9 below.

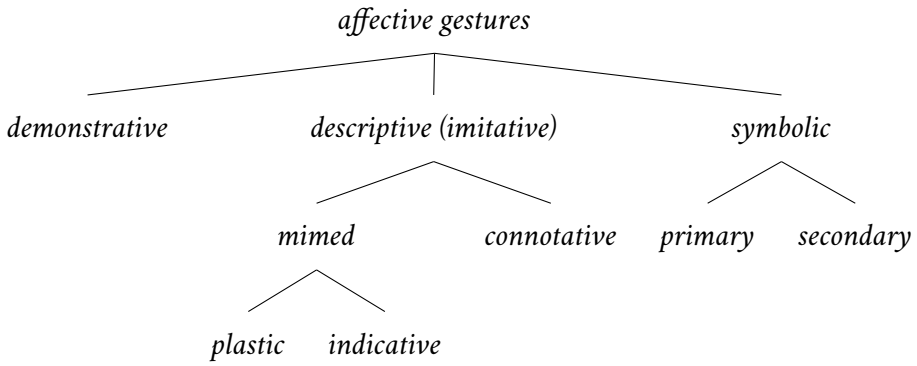


Figure 9: *Wundt's classification of affective gestures*

Apart from neglecting beat gestures, we can see that Wundt recognized the semiotic axes that underlie the modern classification: he distinguishes indexical gestures and gestures representing some semantic content – further differentiating between *descriptive* (iconic) and *symbolic* gestures (the latter corresponding to metaphoric gestures in the modern view). The process of metaphorical extension, a key concept in modern cognitive linguistics, is foreshadowed in what Wundt describes in his definition of symbolic gestures:

> "We may use this kind of gesture if, through some kind of conceptual extension by association, it refers indirectly to the idea it represents […]. Since one thinks of a 'symbol' as a sensory image that is supposed to represent a concept differing from itself but related to it by association, then a 'symbolic gesture' will be, in this general sense of the meaning, one which stimulates a certain sensory image in order to tie together different thoughts associated through inner qualities" (Wundt, 1973, p. 88).

Primary symbolic gestures are those that originally evolved as symbols, whereas secondary symbolic gestures evolved from imitative gestures via a process of gradual

[46] Recall Merleau-Ponty's conception of gesture (Section 3.1).

obscuring of the iconic link. The division of imitative gestures corresponds to the degree of iconicity. In his description of how imitative gestures (of the mimed kind) performed by one may induce the same affect in the other, Wundt had described something, as Pim Levelt aptly pointed out (2013, p. 174) in principle similar to what later became known as the *mirror neuron system* (discussed in more detail in Section 3.2.1).

The last aspect of Wundt's treatise on gestures resonating with the modern views that I want to mention here is the close attention he pays to sign languages of the deaf, and his acknowledgement of their status as natural communication systems. However it may seem, from the today's perspective, that Wundt was "ahead of his time", such a view had been present before Wundt. The suppression and sidelining of sign languages (in the education of the Deaf and in general) started to prevail (for a number of reasons) in the late 19th century, marked by the enforcement of oralism at the infamous Milan Conference in 1880 (Lane, 1992, cf. also Baynton, 1996).

Although there was a considerable disruption in cognitive linguists' interest in gesture between the early 1980s and the end of the 20th century,[47] nowadays CL-based gesture studies represent a thriving research area indeed, fulfilling their early-recognized potential for the explanation of the cognitive underpinnings of language.

Criticizing the then-mainstream paradigm – Generative Grammar (GG) – the aim of CL was to provide a model of language as well as a conceptual system adequate to theories of cognition in general. The key principles of CL offer an alternative view to the GG idea of language as an autonomous and self-contained system:

(i) mental representation of language is not an autonomous (inborn) module of the mind – it is functionally interconnected with other cognitive capacities, and its ontogenetic development relies on learning and adaptive processes distributed across cognitive domains

(ii) nor is language as a system of signs modular – whereas GG presupposes the modularity of grammatical levels with distinct modules for syntax, lexicon, and semantics (the role of syntax being paramount), CL focuses on a common interface for all grammatical levels

(iii) the organization of lexicon is not a hierarchical tree-like structure. Rather, it is based on the extent of prototypicality of category membership, while categories are characterized by their fuzzy boundaries

(iv) the purpose of CL is to account for how language actually works (usage-based approach) – focusing primarily on what GG would set outside the core of linguistic inquiry as performance (GG focuses on the universal underlying principles: *competence*).

[47] This hiatus was in particular caused by the lack of sufficient analytic tools at the time. However, it also has to be noted that there had been a more general tendency to narrow the scope of the research problems among cognitive linguists in this period, leading to addressing a rather limited number of aspects of the theoretical complex of CL and providing poor empirical ground for it, if any at all (see Divjak et al., 2016).

Principles (i) and (iv) are the most significant ones for the inclusion of gesture into the grammatical framework:

Principle (iv) is generally in accord with the interactionist approach. CL takes the default form of language, i.e. everyday, spontaneous, spoken dialogic interactions, into account. However, it does not deal primarily with the problem of how to capture and analyse them. Instead, CL focuses on the *performance* phenomena that were once dismissed as peripheral.

Principle (i) suggests that the generativist idea of an autonomous and universal symbolic system of concepts – the "language of thought" (as it was influentially claimed by Fodor (1976), language-like but ontologically distinct from the linguistic representation – is no longer tenable. Instead, CL favours the view of conceptualization that is grounded in (and phylogenetically derived from) immediate human experience with the lived world and the body through which the individual experiences the world. In other words, conceptualization and human cognition in general are *embodied*. Embodiment thus provides a solution to a paradox inherent in the modular view – referred to as the *symbol grounding problem* (Harnad, 1990): if our conceptual system is based on intrinsic symbols, where did the symbols come from in the first place? As the following section will illustrate, embodiment theories entail a very concrete resolution to this puzzle. The idea of a symbolic system upon which conceptualization is based as an autonomous, language-like code is untenable, since it only enables us to account for its nature and origin in its own terms, thus leading to a tautology. A view of cognition as embodied, i.e. grounded in the general sensory-motor processes and related cognitive skills such as vision, motor control and – crucially – proprioception, not only debunks the universalist stance but also brings in an appropriate argument against behaviourism: we are indeed able to address mental processes while doing so beyond the sphere of sheer speculation. Viewed from this perspective, the turn to embodiment does not truly represent the "second cognitive revolution" as it is sometimes framed (coined by Harré and Gillett, 1994). Rather, labelling the proponents of the embodied view as the "second generation of cognitive science" would be more appropriate, as suggested by Sinha (2007, p. 1266).

3.2.1 Gesture as manifestation of embodied cognition

In the wake of the emergence of CL, the idea of embodied cognition played an essential role in reconsidering the role of gesture in language and cognition. Whereas for the proponents of the nativist view of language, gesture and other alleged "paralinguistic" features have mostly remained without relevance, or as a mere epiphenomenon of language, worthy of attention only in the context of "performance" but not "competence",[48] cognitive linguists acknowledged the importance of gesture as early as in the

[48] See, e.g. Chomsky's remarks on gesture: "[gesture and speech] are in tandem, and some common source is obviously controlling them both; they're just too well correlated for anything else to be the

late 1970s.[49].

The beginning of what might be called "turn to the body"[50] in linguistics has traditionally been associated with Lakoff and Johnson's work (1980) which was on of the first publications to introduce the idea that the conceptual system underlying human language is organized through schemas based on the physical/sensory experience of the real world, and that these embodied schemas work universally across all levels of abstraction of concepts.

Although Lakoff and Johnson did not directly address the question as to how these schemas (*conceptual metaphors*) are visualized through metaphoric gestures, this issue was very soon raised by McNeill and Levy (1982).[51] The authors, as mentioned in Chapter 2, were the first researchers to empirically approach the role of gesture in conceptualization, assuming that gesture may manifest what McNeill (1979) called *concrete models* – metaphorical concretizations of abstract concepts, basically analogous to Lakoff's and Johnson's conceptual metaphors:

> "Some gestures seem to reveal concrete models to direct observation. The form of the gesture conveys information that suggests the existence and character of the concrete model with which the speaker is representing the information he conveys in speech" (McNeill and Levy, 1982, p. 272).

The analysis of gesture use in narratives elicited using a short animated film as a stimulus – a method that was later to become standard in gesture studies – confirmed the assumption, as gesture exhibited its close relation to the meaning of coinciding speech at the level of discourse as well as at the level of individual words.

Despite the early recognition of gesture as a potentially important part of the linguistic form, actual attempts to integrate gesture into the CL framework were not made for almost two decades (barring a gesture-related CL paper by McNeill and Levy (1982)). The main obstacles that caused this gap were eliminated when sufficient technical equipment needed for the study of the visual aspects of language production – video-cameras and later also annotation software – became broadly accessible. Until then, gesture had not been mentioned explicitly in relation to CL very often (cf. Lakoff, 1987).

Wilson (2002) distinguishes several approaches to embodiment representing general tendencies in cognitive sciences. One of them, perhaps a bit misleadingly labelled

---

case. Nevertheless, the system of gestures is very different in its underlying principles from the system of language" (Chomsky et al., 1983, p. 40).

[49] In 1977, Lakoff argued that thought, perception, emotions, cognitive processing, motor activity, and language are all organized in terms of the same kinds of structures ((Lakoff, 1977, p. 246)

[50] Or "somatic revolution" – depending on which framing we prefer: that of a *revolution*, or that of a *turn/generation* within the revolution

[51] McNeill and Levy note that their interest in metaphoric gestures was inspired by Lakoff who, in fact, coined the term *metaphoric gesture* (McNeill and Levy, 1982, p. 274).

"offline cognition is body-based", applies very well to how embodiment has been conceived in the Lakovian tradition. This approach to embodiment is based on the assumption that mental structures initially specialized for interaction with the environment function the same way even without the interactional context:

> "Mental structures that originally evolved for perception or action appear to be co-opted and run "off-line", decoupled from the psychical inputs and outputs that were their original purpose, to assist in thinking and knowing. [...] In general, the function of these sensorimotor resources is to run a simulation of some aspect of the psychical world, as a means of representing information of drawing inferences" (Wilson, 2002, p. 633).

Since the 1990s, immense progress in neuroimaging methods and the input from computer sciences has led to a more fine-grained localisation of the neural centres and networks associated with language production and comprehension, as well as with storage and retrieval from the mental lexicon. This facilitates our understanding of language processing in both healthy and impaired individuals. Development in neurology has also supported the assumptions of the CL approach to embodiment. More light has been shed on the basal linkage between neural structures responsible for speech production and motor control – both localized in Broca's area in the frontal lobe of (in most cases) the left hemisphere.

The mirror-neuron legacy

Another considerably significant impact on the theory of embodied cognition came with the discovery of the so-called mirror neurons in the brain of the pig-tailed macaque (*macaca namestrina*) (di Pellegrino et al., 1992). Mirror neurons are neural structures in area F5 of the macaque brain responsible for manual and oral motor skills that have been observed to activate not only when a motoric action is performed, but also when it is being perceived (performed by other individuals) or visually represented (as an image stimulus or a mimetic gesture). Since area F5 in the macaque is (by the majority of neuroscientists) recognized as a homologue to certain parts of Broca's area in the human brain, an assumption soon arose that the mirror neurons might be the coveted neurological evidence for the representation of action – a basis of embodied cognition (e. g. Rizzolatti and Arbib, 1998). This finding inspired hypothesizing about the origin of the human conceptual system, among them e.g. Gallese and Lakoff's (2005) proposal assuming the direct embodiment of the conceptual structure via certain neuronal networks. Premotor and parietal neural structures including mirror-neurons comprise specialized "functional clusters" – bi-directional ("multimodal", i.e. serving for performing as well as imaging an action) neural networks – a direct source of the conceptual representation of concrete sensory-motor actions. Crucially, this system also

embodies more abstract concepts, metaphorically derived from the physical experience. Abstract reasoning thus "*exploits* the sensory-motor system" (Gallese and Lakoff, 2005, p. 473, emphasis in the original).

Such a strong version of embodiment is not accepted universally. For instance, Arbib (e.g. 2014) holds a more moderate position, arguing that embodiment has a graded character and that the conceptual system in its entirety cannot be explained in terms of embodiment. However, from the evolutionary perspective, human communication evolved from a fully embodied primordial stage. According to Arbib's model called *Mirror System Hypothesis* (Arbib, 2013; Arbib et al., 2014), the emergence of human language was catalysed by mirror neurons underlying a general (i.e. not exclusively human or primate) visual system for motion recognition and action. On top of this system, an analogous mechanism evolved (as a part of what Arbib calls "language-ready brain",[52] thus allowing humans to map semantic information onto the real-world experience and to communicate those mappings with others in the form of conventionalized form-meaning combinations. Interconnected with the communication mechanism is the so-called *schema network* – an embodied conceptual system based on the action-perception feedback loop: "our internal states […] shape how we act, and these actions then shape what we perceive, updating that internal state into a process and so the cycle continues" (Arbib, 2013, p. 109).

Theories based on the mirror system are not, however, accepted generally among neuroscientists. For instance, Hickok (2014) aims his criticism at – among other things – the assumption that the mirror neuron system is also involved in language comprehension, since there is, according to Hickok, no basis for this claim in terms of localization: Broca's area is associated with language production, not comprehension.

Nevertheless, there seems to be a sufficient consensus among neuroscientists that the core of the mirror-neuron based research is valid and that significant discoveries are yet to come (cf. discussion in a special issue of *Language and Cognition* from 2015 including a reaction to Hickok's criticism by David Kemmerer (2015)).

The mirror-neuron hypothesis provides some support for the theories of language evolution and the role of gesture therein. Most notably, Stephen Levinson and Judith Holler (2014) offer an alternative view to the "gesture-first" hypotheses (Corballis, 2002; Hewes, 1973), assuming that gesture was an evolutionary precursor of vocal language. According to Levinson and Holler, the system of human communication (or as they call it, *interaction engine*) emerged via co-evolution of manual and oral communication. In other words, human communication has always been multimodal – only the dominant channels changed. The primary role of vocal language in modern humans is only a result of physiological changes such as the evolution of voluntary control of breathing and the descent of the larynx.

[52] See also Levinson's and Holler's *interaction device* below and McNeill's concept called *Mead's loop*, "[an] adaptation in the evolution of humans, wherein mirror neurons were 'twisted' to respond to one's own gestures, as if they were from someone else" (McNeill, 2012, p. 70).

Challenging the idea of non-perceptual nature of human conceptualization, then widespread in cognitive psychology, Lawrence Barsalou formulated a theory (1999) assuming a simulation-based mechanism underlying the conceptual system, including abstract concepts. Since its advent, this theory has resonated in cognitive sciences, and it has inspired a major research direction in cognitivist gesture studies as well.

In Barsalou's theory, the basic unit of the conceptual system, i.e. concept, is viewed as a *simulator*. The simulator itself represents a sub-system of the so-called *perceptual symbols*, unconscious representations of perceptual experiences, existing purely at the neural level as "records of neural states that underlie perception" (p. 582). Perceptual symbols provide only very coarse-grained information about perceived reality combining all sensory sources (not only visual, such as colour, texture or shape but also olfactory or aural). They are not stored as discrete units but are subject to perceptual adjustment (re-framing). A combination of related perceptual symbols constitutes simulations of events or items that are not directly perceived. These simulations are based on selected traits of the given event or item, always framing only its certain aspects, in a more or less distorted way.

In line with the assumptions of CL, language is part of the general conceptual system, being a system of simulations ontologically equal to any other kind of simulations. However, linguistic simulations play a unique role in the human conceptual system, as they are mapped onto non-linguistic simulation, providing the entire system with a structure as well as clues for the generation of novel simulations: "[o]nce simulations for words become linked to simulations for concepts they can control simulations. […] Linguistic symbols index and control simulations to provide humans with a conceptual ability that is probably the most powerful of any species" (1999, p. 592). This indexing and control work inter-subjectively, ensuring a productive re-invention of the simulator of an individual subject to another and thus enabling sharing knowledge and establishing common ground, i.e. the cognitive abilities conditioned on a key feature of the human language: displaced reference.

Abstract concepts are also based on simulation. Abstract simulators draw on generalizations over a multiplicity of concrete events experienced in time and situated in a specific background – depending greatly on perceptive information provided by introspection (Barsalou and Wiemer-Hastings, 2005).

Implications for the origin and nature of gesture offered by both mirror-neuron- and simulation-based views of embodied cognition are basically alike.[53] If the conceptual system is based on the enactment of perceived actions, a possible explanation of the very phenomenon of gesture lends itself: gesture is, or at some point in the evolu-

[53] It is not surprising that mirror-neurons have been embraced by many proponents of mental simulation theories of human cognition as "neurological evidence" in their favour (e.g., Fischer and Zwaan, 2008; Gibbs, 2006).

tion of human communication it was a visual extension of the enacted action, serving either as a facilitatory tool or reflecting the cognitive processes occurring during conceptualization, or both.

Hostetter and Alibali (2008) present the *Gesture as Simulated Action* (GSA) model, which attempts to capture the inner workings of the cognitive mechanisms behind gesture production. According to GSA, the potential for gesture realization emerges at the neural level when neural structures responsible for simulation are activated, as those structures are located in the same area as neural structures associated with motor control. The actual realization of gesture then depends on the following factors (pp. 503ff):

First, one must consider the *strength of simulation activation* as well as the *strength of the motoric activation*. When the activation of the simulation areas is not strong enough, it does not trigger the activation of the adjacent motor-control neural structures. Simulations differ in their motoric affordance, as some events involve movement more directly than others. This does not correlate with the degree of "abstractness" of a given concept, because abstract concepts are – via metaphoric extension – processed in the same way as concrete ones, which is manifested in metaphoric gestures (2.2).

Second, the gesture realization in the GSA model depends on what Hostetter and Alibali call the *height of gesture threshold*. Whether the potential gesture is realized or not results from various subjective and situational conditions. For instance, there is individual variation in how simultaneous activation of the simulation- and gesture-production related areas occurs. Apart from that, subjective cognitive abilities need to be considered. As the authors point out, it actually requires more cognitive effort to suppress gesture production – the simultaneous activation is, of course, an automatic process, but can be controlled consciously. Hostetter and Alibali mention experimental evidence showing that when carrying out complex cognitive tasks, subjects produce more gestures than when they can focus only on their own production. Another factor influencing the likelihood of gesture realization is the immediate context of communication – gesture production of other speakers, as well as pragmatic and social constraints that apply in a given situation.

Although GSA provides a very elegant schematization of gesture production, it is limited only to representational gestures, i.e. the "movements that represent the content of speech" (p. 495), including deictics. The way referential gestures, such as simple pointing, are generated as a by-product of mental simulation is not, however, accounted for by the authors. Their theory is thus best suitable for iconic and metaphoric gestures (which should be, as discussed in Section 2.2, considered a special kind of iconic gestures). We can see that the McNeillean categories are viewed here as discrete units, although in actual language interaction the boundaries between gesture functions are often obscured. A truly adequate model should take this multifunctionality into account, proposing a mechanism that would also apply for beats

and deictics: showing how the realization of other functions of gestures (apart from representation of conceptual content) fits in the model's schematization, or whether it requires a separate cognitive model. This is true for perception as well: GSA does not address the role of simulation in the perception of gestures and with respect to language comprehension.

A considerable amount of work in the cognitively oriented studies of gesture has been carried out with a focus on gesture as the manifestation of *image schemata*. Recalling the definition of the image schema (Johnson, 1987, p. xiv) introduced above, we may look at it as a form of embodiment at a more rudimentary level than conceptual metaphors that are built upon them. Image schemata are characterised as "preconceptual and non-propositional" (*ibid*., xvi). As far as the degree of abstraction and the evolutionary perspective go, gestures may be situated between *perceptual symbols* and *mental representations* as they are generally understood in CL. See Section 2.2 for more detail on gestural representation of image schemata.

I will follow-up on the discussion of the psycholinguistic models of gesture production (and comprehension) in Section 3.2.2.

Bodily relativity

Another noteworthy line of research that has brought evidence in support of the strong view of embodiment is represented by the study of the so-called *bodily relativity*. This thesis, termed with explicit reference to linguistic relativity, is based on the "bodily-specificity-hypothesis" (Casasanto and Chrysikou, 2011; Casasanto, 2016): if mental representation – including abstract concepts – is embodied, individuals with different bodies necessarily conceptualize differently. A lexical decision study was conducted by Willems et al. (2011) who presented 20 right-handed subjects with verbs of manual action, verbs of non-manual action and pseudoverbs. During the lexical decision task, subjects' motor cortex responsible for the manual motor control of either the dominant or non-dominant hand (two experimental conditions) was stimulated using *Theta Burst repetitive Transcranial Stimulation* (rTMS)[54]. The stimulation affected subjects' recognition of manual action verbs when it was targeted to the right-hand motor cortex, whereas there was no effect with other verbs regardless of the stimulated area. This showed that the neural structures responsible for manual action also play a direct role in the comprehension of the linguistic representation of the given manual action. In other words, executing manual action and understanding manual action at least partly involves the same neural structures: "in this sense, people with different bodies understand the same verbs to mean something different" (Casasanto, 2014, p. 110).

A direct embodiment of the concrete, manual-action related concepts does not

[54] A method based on emission of electro-magnetic waves targeting specific parts in the cortex for a brief moment to evoke short-term excitation within a given area (Hallett, 2007).

necessarily lead to the idea of strong embodiment, i.e. somatic grounding of entire conceptual systems, including abstract concepts. However, results of behavioural studies speak in favour of the strong view. Casasanto (2009) reports the results of a series of experiments focused on the association between spatial location and emotional valence.[55] American as well as Dutch subjects underwent tests aiming at attributing positive or negative evaluation to stimuli (with neutral emotional value) presented on the subjects' right- or left-hand side. The experiment revealed a systematic association of positive emotional value with the right-hand side and negative with the left-hand side in right-handers. Given the metaphor RIGHT IS GOOD/LEFT IS BAD conventional for both Americans and the Dutch, this is nothing but expectable. What is striking, however, is the finding that left-handers exhibit the very opposite mapping between the left-hand side and positive value, and vice versa.

Moreover, Casasanto and Chrysikou (2011) showed that when hand dominance changes, conceptual re-association of emotional values and spatial locations occurs. First, they studied subjects with dominant hand paralysis after a stroke, who exhibited a remapping of the positive value on the healthy hand in tasks similar to the previous study. This finding was later supported by the results of the same test conducted with healthy subjects who had their dominant hand temporarily incapacitated.[56] Even a very short forced imbalance of hand-lateralization influenced subjects' performance in the behavioural task.

Fringe areas: Insights from language pathology

Studies with subjects with brain damage represent a specific body of evidence regarding the linkage between the neural structures responsible for motor control and the parts of the brain associated with language processing, as well as the role of gesture within this interface. Most of the findings are based on the research of aphasic populations. This area has its peculiarities, providing fascinating insights into the physiological basis of language but often with limited potential for generalization as they are typically based on case studies involving aphasic patients with unique conditions. Nevertheless, there are some general lines of evidence that point in the same direction as what has been summarized here so far.

Based on the neurological affinity between motor control and language production, specific patterns can be expected. Lesions in Broca's area (in the inferior frontal gyrus within premotor cortex in the left[57] hemisphere) lead to impairment of speech production (Broca's aphasia, also referred to as "non-fluent", "anterior" or "motoric") as well as to motoric impairment. Gesture should therefore be affected as manual move-

[55] These tests usually involve presentation of stimuli with positive/negative emotional value in a divided visual field, or attribution of positive/negative values to neutral stimuli presented in a divided visual field.

[56] They were asked to wear a ski glove.

[57] Depending on the lateralization of the individual brain.

ment, in general, is affected: the aphasic patients with Broca's aphasia often have hemiparesis (weakness of half of the body) or suffer from manual apraxia (motoric disorder).

However, in some cases, the capacity to produce gesture is retained. In aphasics with impaired speech production, compensatory use of gesture was reported (Dipper et al., 2015; Kemmerer et al., 2007), manifested in increased frequency of gesture production. The increased use of gesture not only compensates for what is missing in production for the sake of clarity of expression, but it also facilitates lexical retrieval during the speech production itself. Consider a typical gesture occurring during lexical retrieval frequent in neurotypical population (Feyereisen, 2006; Krauss et al., 2000): a repetitive cyclic motion of a hand. In their respective case studies focused on the gestural and verbal expression of motion events, Kemmerer et al. (2007) and Dipper et al. (2015) report (independently) relatively intact gesticulation, exhibiting the same patterns of encoding motion event structure as in healthy speakers of the given language.

On the other hand, in case of Wernicke's aphasia (also "fluent", "posterior" or "receptive") , production is intact in terms of articulation but lacks semantic coherence, co-occurring with a varying degree of comprehension impairment. This class of aphasias does not necessarily involve motoric impairment, as the lesions are localized in Wernicke's area (in superior temporal gyrus of the (mostly) left hemisphere). Sekine et al. (2013) analyzed gesture production of a relatively large number of aphasic subjects (46), using data from a corpus of aphasic speech,[58] including both Broca's and Wernicke's cases. To assess gesture usage, the authors measured the frequency of gestures and assigned them to 12 gesture types (combining pre-existing typologies and inventing new types), completely ignoring their potential multifunctionality. Results of the analysis revealed expectable patterns: in general, aphasic subjects (compared to a neurotypical control group) produced more gestures; weighted against the total amount of speech produced, Broca's patients produced significantly more gestures than Wernicke's. The former group was reported to produce more "semantically coherent" gestures linked to the content of speech and serving as the expressive extension of their limited speech abilities, whereas subjects from the latter group produced "a low number of semantically rich emblem, iconic and pantomime gestures and high number of beats and metaphoric gestures which are less communicatively meaningful" (p. 1043), thus showing a significant correlation between aphasia type and type of gesture. Despite its questionable approach to gesture categorization, the general findings of this study support the idea of a common production interface between speech and gesture, inseparable at every level of the production process, from its start at the conceptual level (lexical/constructional retrieval) to articulation. Also, as will be discussed below, an analogical interconnection between gestural and verbal content of communication applies to the comprehension process too.

Extreme linguistic phenomena like aphasias provide highly valuable insights

[58] *AphasiaBank*, MacWhinney et al., 2011.

into the inner workings of neurological mechanisms of language, the kind of insights that could hardly be acquired by the study of a healthy population. I will conclude this section by three other examples of this kind, which in some respects shed even more light onto the intricate interconnectedness between vocal and gestural expression.

Shaun Gallagher mentions cases of congenitally armless subjects who experienced "phantom movements" of arms when speaking – explicitly referring to their own experience of phantom limbs moving as "gesturing", but not when walking (Gallagher, 2005, p. 120).

Another curious case mentioned by Gallagher is a man (IW) who lost the capacity of proprioception from the neck down following an unspecified autoimmune disease in his adult age. Thoroughly studied by physicians as well as psychologists, IW was subjected to an experiment (p. 112ff) designed to test his ability to produce co-speech gesture. IW himself reported to deliberately abstain from using gestures, as he was not able to control his manual action properly. In the experiment, a special blind was put in front of him, blocking his view of those parts of his body that were affected by the loss of proprioception. When asked to retell a cartoon he had watched, IW produced gestures in a normal way, performing beat and iconic gestures synchronized with the related verbal content – completely unaware of what he was doing.

*

* *

From the neurological perspective, talking and gesturing are two sides of the same coin. In this section, I presented a digest of the evidence of neural interface between how spoken language and accompanying gestures are processed, supporting the idea of embodiment of human cognitive capacities, including language. In the following section, I will approach this issue from a psycholinguistic perspective, reviewing models that were proposed to describe the cognitive mechanisms involved in cognitive processing of speech and gesture.

3.2.2 Psycholinguistic models: co-speech gestures' role in language processing

Speech and gesture are intertwined at the level of conceptualization as well as articulation. The question that stands out is what happens in between. Between the activation of neural circuitry to which we attribute conceptualization and the physical realization of visible hand movements and audible sounds of language lies a dark area: a testing ground for psycholinguistic hypotheses about language processing.

In psycholinguistics, a processing model usually schematizes the communication process or certain parts of it (typically the production of spoken language) via the metaphorical framework of processing units – functional components

based on our knowledge of how general cognitive abilities such as memory or perception work and information flow between specific nodes connecting individual components.

But not all the accounts reviewed here are *models* in this technical sense, but rather sets of assumptions about the nature of speech-gesture processing. Some have already been mentioned in the previous section. Below, I will look more closely at some other theoretical models – but it will by no means be a comprehensive summary.[59] The six models discussed here, and the GSA model addressed in the previous section, are the most relevant for this study. It was these models that were used in previous research to base hypotheses or support interpretation of the findings, and their adequacy will be discussed again in the context of results of the present study.

Several models of language processing are based on a model developed by Pim Levelt (1989, with many later adjustments and reformulations). The core idea behind Levelt's model is the hierarchical activation of separate modules (or *strata*): beginning with the conceptualization level (conceptualizer) at which conceptual content is turned into a "preverbal message", followed by activation of a corresponding lemma in the mental lexicon (formulator) that outputs selected phonological and grammatical structures for final articulation (articulator). In later elaborations, Levelt's model also captures comprehension, treating it as a reverse process: the stream of perceived speech is processed in a parsing module hierarchically parallel to the formulator, and its output is fed back to the conceptualizer. Schematization of the basal nodes of the architecture of Levelt's processing model is captured in the diagram below (Figure 10).[60]

Interface Model and Information Packing Hypothesis

The model proposed by Sotaro Kita and Aslı Özyürek (2003) is an extension of Levelt's model of production. The interface model is based on an assumption (the Interface Hypothesis) supported by the findings of empirical studies of the verbal and gestural expression of motion event structure in different languages (thoroughly discussed in Section 4.2.1). The research revealed systematic correspondences between coding of the motion event structure in speech and in gesture, reflecting cross-linguistic differences between the conceptualization of the motion domain. According to the *Interface Hypothesis*, which follows the idea of "weak" linguistic relativity,[61] i.e. the "adjustment of one's thought to the vast idiosyncrasy of the lexicon […] performed on-line at the moment of speaking" (p. 27), the form of gesture[62] originates at the level of Levelt's

[59] For models omitted here, see reviews in Feyereisen (2013) or a representative collective monograph on gesture production and comprehension edited by Kelly, Church, and Alibali (2017).

[60] The diagram was drawn using Emiel van Miltenburg's code (https://gist.github.com/evanmiltenburg/9f1202391d6f9e90fa531a99970eec7e).

[61] Slobin's *thinking for speaking* – see Section 4.2.1 for more detail.

[62] That is, only representational gestures.

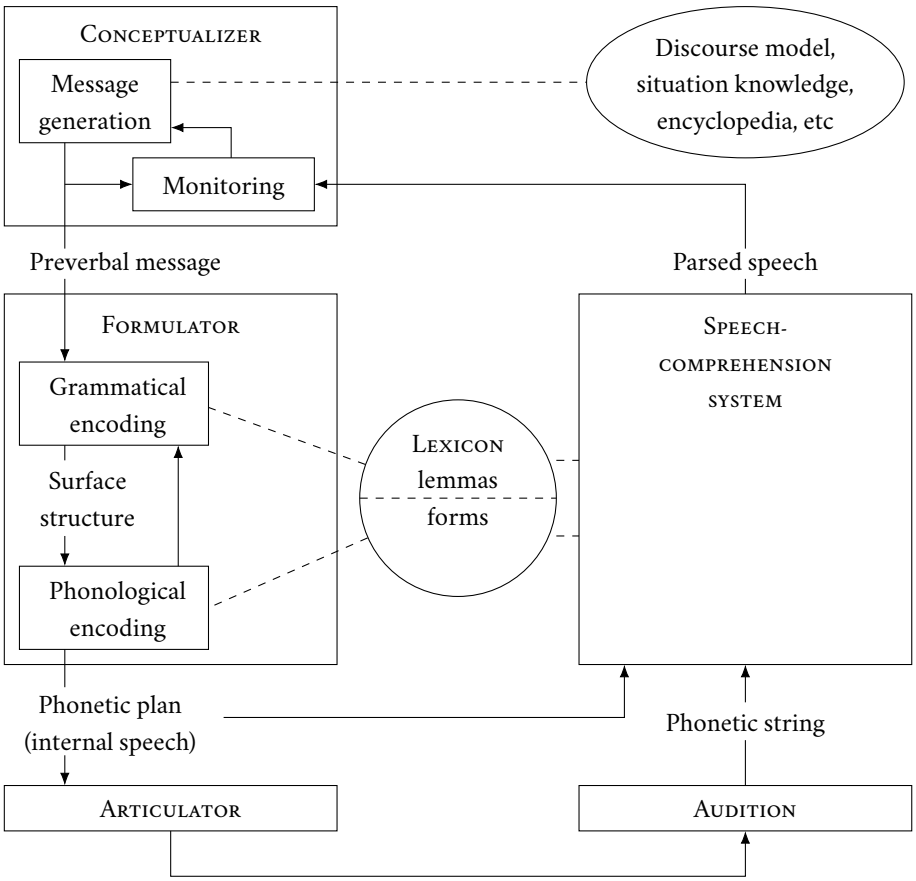


Figure 10: *Levelt's model.*

conceptualizer. At the conceptualizer level, there is a deeper hierarchical structure that consists of the communication planner responsible for the generation of the "raw" content and its assignment to the spatio-visual (*action generator*) and spoken (*message generator*) modalities. The action generator and message generator function as an interface, allowing for bi-directional adjustment of both gestural and verbal expression during production. Figure 11[63] depicts the model's mechanism.

Importantly, the model does not presuppose that the form of the gesture is determined by the communication planner, as it is to a certain degree shaped by the spatio-motoric features of the referent that are stored in the working memory and are not (or do not have to be) communicated verbally. This dual nature of representational gesture generation makes the Interface Model generally compatible with McNeill's Growth Point model (3.2.2) as well. Generally speaking, the Interface Model is also compatible with GSA and theories of embodied cognition in general, as it assumes that gesture is a product of action of schemata which drive the action generator. But similarly to the GSA model, the Interface Model limits itself only to iconic and metaphoric gestures: there are no implications as to whether we should assume a completely different cognitive mechanism for the generation of pragmatic and deictic gestures, or a similar but

[63] Source of the diagram: Kita and Özyürek, 2003, p. 28.

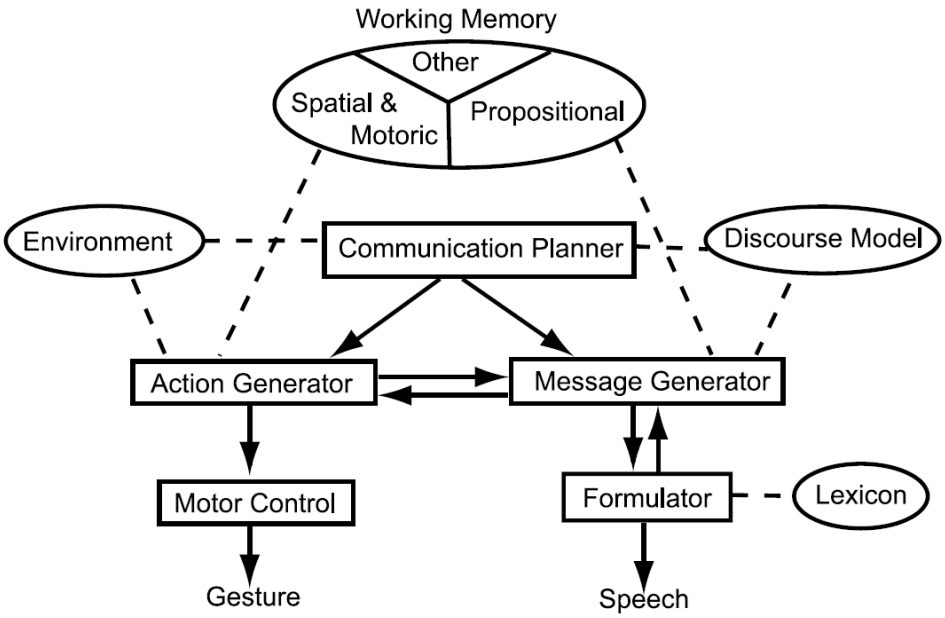


Figure 11: *The Interface Model*

separate one.

As a general sketch, the Interface Model not only offers a possible mechanism that underlies the production of speech and gesture, but it also hints at the explanation of the "relativistic" effects in non-linguistic domains, provided that analogous mechanisms apply to the possible "language comprehension – visual perception interface". However, Kita and Özyürek do not address the comprehension perspective. What the "full picture" might look like – that is, Levelt's complete framework with both production and comprehension processes and verbal as well as gestural component – is captured in Feyeireisen's synthesis of several proposed models (Figure 12).[64]

Kita and Özyürek's model is supported by a considerable body of evidence from the motion domain. However, the evidence is limited to a specific type of iconic gestures expressing simple motion events. Moreover, these gestures were mostly captured in controlled conditions during narrative tasks designed to elicit exactly those kinds of gestures.

Nevertheless, the Interface model is a relatively open framework with a potential for further elaboration, which is perhaps more crucial for it than additional empirical grounding. As de Ruiter puts it (2017, p. 66), the Interface Model is empirically unfalsifiable as it does not enable formulating specific hypotheses about gesture – speech relations.

Closely linked to the Interface Model is the *Information Packaging Hypothesis* (IPH). Chronologically, the IPH preceded the Interface Model – it was first put forth by Kita (2000), but it was later incorporated into the Interface Model architecture

[64]Source of the diagram: Feyereisen, 2013, p. 158.

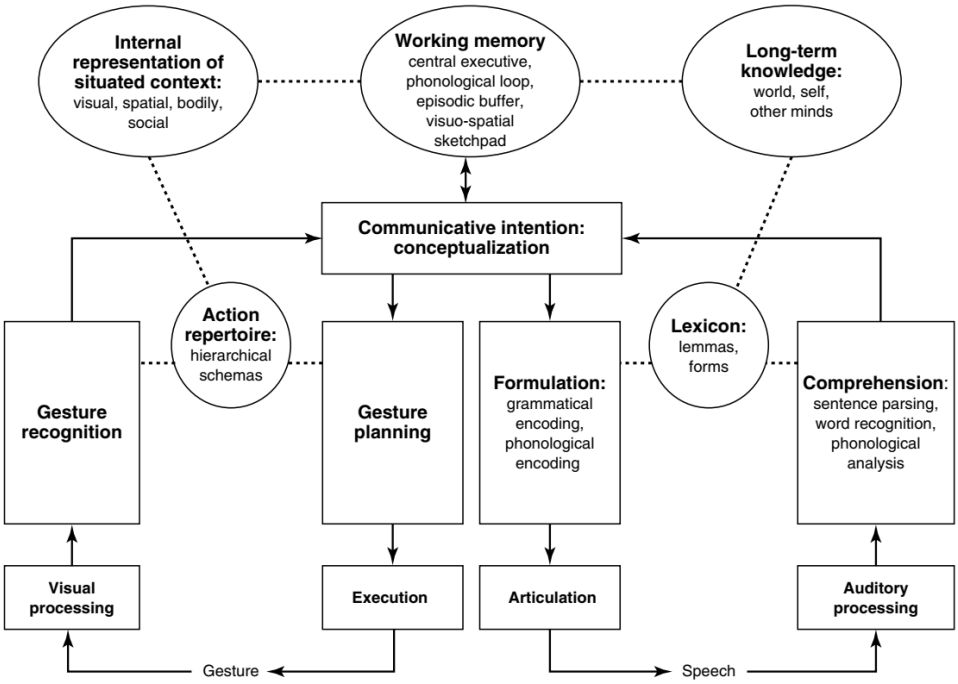


Figure 12: *Feyereisen's synthetic schematization of gesture-enriched Levelt's model*

(Alibali et al., 2017). The IPH is also based on the dual nature of representation – referred to as "analytic thinking" (which is the primary mode of information organization in speech) and "spatio-motoric thinking". Both modes of thinking contribute to how information is structured in speech: "the collaboration between the two modes provides speakers with wider possibilities to organize thought in ways suitable for linguistic expression" (Kita, 2000, p. 180).

Sketch Model and Asymmetric Redundancy Hypothesis

Another psycholinguistic model that builds upon Levelt's framework is the *Sketch Model* and its recently modified version – the *Asymmetric Redundancy model* (AR), both formulated by J. P. de Ruiter (2000; 2017). The Sketch Model accommodates Levelt's architecture and McNeill's idea of (representational) gesture as expressing different aspects of the mental content than those expressed in the accompanied speech. This model is sometimes called the "Postcard Model" following the metaphor of a postcard that carries one communicative intention split into the "imagistic" (gesture) and "propositional" (speech) sides. If we take a look back at Figure 12 and for once we ignore the comprehension/perception parts of the schema, we are effectively left with the Sketch Model. Here the *Action repertoire* is called the *Gestuary* – a gestural analogy of mental lexicon, where the schematic templates for gestures are stored. Thus, it is the *Conceptualizer* where gestures are generated, and there is no feedback from the verbal channel in further stages (i.e. the Formulator and its gestural counterpart at the same

hierarchical level, the Gesture Planner, are separated).

Later, de Ruiter abandoned this idea of complementary but discrete channels, acknowledging the evidence that had meanwhile accumulated,[65] showing that representational gestures actually tend to reflect the structure of the verbal expression. This co-expressiveness of gesture is not exhibited only at the lexical level (which is the scope of the Interface Model), but also at the level of complex syntactic constructions and even at discourse level. To account for both gesture-speech interface and its variable scope, de Ruiter introduced the *Asymmetric Redundancy Model* (2017). It assumes redundancy in the sense of duplicity of gestural and verbal expression, both generated alike in the Conceptualizer which contains the subordinated Gesture Planner responsible for further gesture generation. The information that goes from the Conceptualizer to the Gesture Planner is fully constrained by the information that the Conceptualizer sends to the Formulator (see Figure 13).[66]

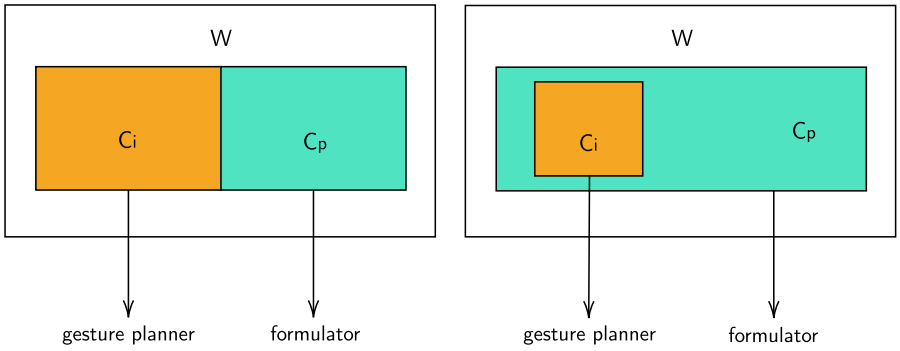


Figure 13: *Sketch Model and AR-Sketch Model*

The asymmetry within the model is twofold. First, the model is asymmetrical (compared to the original Sketch Model) because it assumes subordination of the Gesture Planner to the Formulator. Second, the model assumes that the Formulator exploits the whole representational pool of the Conceptualizer, including the imagistic representations, whereas the Gesture Planner makes use only of the imagistic (i.e. spatio-visual) content. Unlike the Interface Model which is based on a feedback loop between the Action Generator and the Message Generator, the AR model proposes asymmetric constraining of gesture expression by the verbal part of the utterance.

The AR model does not offer an explanation of the comprehension/perception processes. On the other hand, de Ruiter goes far enough to suggest the possibility of incorporating beat gestures. According to the AR model, the gesture form is constrained by the aspects of communicative intention selected by the Formulator, but as de Ruiter

[65] Examples of this kind of evidence will be given in Sections 4.2.1 and 4.2.2.

[66] Diagram adapted from de Ruiter, 2017, p. 68 (W = working memory, C_i = imagistic representations, C_p = propositional representations).

points out, the pre-selected features do not have to be purely visuo-spatial. In the case of beat gestures, it is the information about rhythm that is coded: even though they do not carry any representational content, beat gestures "represent a gesturally expressed rhythmical *signal*" (p. 70, emphasis in the original). However, the author also admits that "[p]erhaps the rhythmical aspects that are signalled by beats are to be found at a higher level of representation" (*ibid.*).

The Growth Point Hypothesis

We have seen that McNeill's influential theoretical model inspired or somehow influenced all of the approaches reviewed above. The reason why I address the *Growth Point* hypothesis last is that McNeill's stand is in direct opposition to the view based on the information process metaphor with its modular architecture (Feyereisen, 2013). Instead of a model in the technical sense, McNeill builds his account of a mechanism underlying production of both speech and gesture, upon what he calls (after Vygotsky, 1986) a "minimal psychological unit": the Growth Point. It is the most basic cognitive unit in which the imagistic (gestural) and linguistic (verbal) components constitute a dynamic equilibrium between the two opposing semiotic modes. McNeill defines the Growth Points (GP) as a "minimal unit of an imagery-language dialectic" (2005, p. 105), "cognitive package that combines semiotically opposite linguistic categorial and imagistic components" (2013, p. 135), the "smallest package of gesture speech unity" or "[m]inimal packages of language embodiment" (2017, p. 80).

What is the nature of this dialectics? Let us recall the axes of the Kendon-McNeill's continuum capturing different dimensions of the gesture-speech relationship (Section 2.1), specifically, the fourth dimension that regards the "character of semiosis". According to McNeill, gestures are global (i.e. non-compositional) and synthetic (i.e. a single meaning mapped onto a single unit), whereas linguistic signs are segmentable and analytic. Thus, the two semiotic modes in which the imagistic and linguistic components exist create a dynamism that gives rise to an expansion – the process of expression (hence the metaphor of "growth"). From the linear speech perspective, Growth Points occur when a need for "mark[ing] of a significant departure in the immediate context" (p. 86) emerges. This is the reason why McNeill borrows the Vygotskian term "psychological predicate" here – echoing also the notion of communicative dynamism (Firbas, 1992). The process of growth (also called "unpacking") ends when stability is achieved: the expressive potential of the Growth Point is realized in the form of a linguistic construction.[67]

The key property of the GP hypothesis is the co-expressivity of gesture and speech, which stems from their opposing semiotic status: they are co-expressive as they express the same conceptual content, but their different semiotic nature leads to a different framing of the same event via gestural and verbal expression.

[67] McNeill uses this term with an explicit reference to *Construction Grammar* (Section 3.2.3).

McNeill gives an example of a person describing a simple motion event in an elicitation task.[68] When retelling a part of a cartoon in which one character climbs up a drainpipe on the inside, the narrator accompanied the spoken expression *goes up through the pipe* with an iconic gesture using a cup-like handshape moving upwards with the stroke of the gesture coinciding with "up through". McNeill argues that what must be verbally expressed via separate units that combine into the target meaning is expressed by the gesture in a global and synthetic way, integrating all the aspect of the event at once and non-compositionally.

The Sketch Model and the Interface Model share this idea of duality of representation in the core of the process of generation of multimodal utterances. But what might at first sight seem as a theoretical affinity is in fact more complicated. McNeill's model does not presuppose any kind of hierarchical structure. In fact, the production mechanism is in this case conceptualized within a completely different metaphorical frame. Whereas hierarchical models, in line with conventional ways of imagining processual nature of various mental processes in cognitive psychology, are built upon a top-down structure, GP is defined as a linear flow: a continuous, sine-wave-like restoration of representational entropy, so to say.

McNeill himself likened GP to "thinking for speaking" (2005, p. 150 and *passim*), i.e. processes described by Slobin (1996) as language-specific cognitive processing activated on-line during language production – a weak reformulation of the so-called linguistics relativity hypothesis (Whorf, 1956). In the hierarchical perspective, thinking for speaking occurs at lower levels: we saw that in the Interface Model, it corresponds to the feedback loop between the Formulator, the Message Generator and the Action Generator. McNeill's model, however, does not rely on higher level modules to provide input for lower level modules.

The GP dynamism is context-dependent, and this situatedness in the external world represent another aspect which distinguishes McNeill's view from the hierarchical models: "[i]n the dialectic model, [unlike in the information processing frameworks], context is not 'input', it is an essential component of thinking for and while speaking, inseparable from the process itself" (p. 86).

As I pointed out above, McNeill often invokes Merleau-Ponty. McNeill's approach to the cognitive basis of multimodal communication relies as much as possible on external phenomena: situatedness of speakers in the immediate context and intersubjective experience, and he frequently borrows phenomenological terminology using notions such as "dwelling" or "inhabiting" to characterise meaning-making processes related to the use of gesture.

Unlike the hierarchical models addressed above, the scope of the GP theory reaches beyond the lexical level. First, understood as dynamic changes of representational entropy, the notion of GP is only meaningful at the level of discourse. Sec-

[68] The same as was for the time used by McNeill and Levy (1982) and which is described in Section 4.2.2.

ond, the multimodal expression born out of GP often embodies the given event as a whole, with a single polyfunctional gesture but analytically more complex verbal affiliate, given the synthetic – analytic binarity. This points to the notion of grammatical construction as it is treated in Construction Grammar: a holistic, complex but non-compositional form-meaning pairing, cognitively stored and processed as a unit (see Section 3.2.3). McNeill in fact points out that the linguistic output of the unpacking process corresponds to the construction in the Construction Grammar sense (2005, p. 82), but he does not comment on this further. Neither does he claim that the linguistic and gestural expression together could be viewed as a grammatical construction, an idea that will be discussed here in more detail in further sections.

McNeill's theory is again primarily concerned with the production part of the communication process. Besides this limitation, it is the central idea of semiotic polarity that has been questioned. Let us reconsider the above example (*goes up through the pipe*). Gesture and speech express the same aspects of the event: the upward movement (*go up*) and its path (*through*). In gesture, both aspects are simultaneously present, while the speech analytically segments the two aspects into a linear sequence. But is it really a "dialectical" relation, when the speaker accompanies through with a gesture that represents both through and goes up? Rather, it seems that the gesture provides an additional focus on the linguistic representation of the event, pertaining to what has already been expressed in speech. Disagreeing with McNeill's interpretation of the example, Adam Kendon offers a different view:

> "the actor is continually adjusting his expressive resources in relation to one another as he seeks to create an "utterance object" that meets his rhetorical aims within the frame of whatever interactional moment he is faced with. I do not see a dialectical struggle, but an orchestration of resources under the guidance of a communicative aim" (Kendon, 2013, pp. 21–22).

The binarity is only notional (analytical vs synthetic encoding) – there is no reason to assume that the addressee perceives the event as represented in the narrator's multimodal expression this way. Indeed, the empirical research of gesture in relation to language comprehension does not support the assumption that semiotic dialectics, regarded as central to production, should play any significant role in the comprehension process (see below).

However seemingly central to the GP hypothesis, the idea of *dialectics* between the two modalities has not been central in the psycholinguistically oriented gesture studies inspired by McNeill's model. Instead, those who adhere to the GP hypothesis or acknowledge its theoretical basis often highlight the *co-expressiveness* or *complementarity* of the gestural and spoken modalities, leaving aside the idea of them being in opposition.[69]

[69] In fact, the body of McNeill's works suggests that, despite the connotations of *dialectics*, for his view of the two representational modalities, too, complementarity is more central than opposition.

*

* *

The models reviewed so far addressed only (or primarily) production. The question remains whether it is possible to account for production while ignoring comprehension. The metaphor of information processing, first put forth by Karl Bühler in his *Organon Model* of communication (1934) but since then deeply rooted in linguistics and beyond, leads to a conceptualization of communication as a dual process, with speakers and addressees as separate "central processing units" (as Jürgen Streeck puts it) between whom the "message transfer" takes place. Even though the modular view of grammar has been rejected by cognitive approaches, modular architecture still prevails in psycholinguistic models of language processing (the very models that are supposed to test and integrate theoretical assumptions of CL).

Also, all of the production models I reviewed are concerned exclusively with one particular functional dimension of gesture. Again, a question should be addressed: can we isolate iconic (or representational) gestures when we talk about cognitive processing of multimodal information? In cases like the one captured in McNeill's example, i.e. in elicited narratives based on simple cartoons, one can indeed observe clear-cut iconic gestures that represent events with high affordance for transparent iconicity (such as simple motion events). But outside the laboratory, in everyday spontaneous interaction, individual functional dimensions of gesture become less distinguishable, and therefore more comprehensible models are desirable, capturing the entire range of how gesture may function in production as well as comprehension of language.

Integrated Systems Hypothesis

Compared to gesture and speech production, considerably less attention has been paid to the comprehension part of language processing. Some implications for comprehension are entailed in the above-reviewed models of production. Here, I will briefly review one approach to gesture-speech comprehension that deserves attention. Further, in Section 6.1, I will review some experimental evidence in this area, focusing particularly on perception of iconicity in multimodal communication.

Spencer Kelly proposes a unifying approach to the role of gesture in language comprehension, which he calls the *Integrated Systems Hypothesis* (ISH; Kelly, 2017; Kelly et al., 2010). Similarly to the production models discussed above, the ISH also stems from the McNeillean theoretical frame. Its basic assumption is that in any given multimodal expression, gestural and linguistic modalities provide mutual focusing. The integration of the two modalities is not limited only to the semantics of the multimodal expression, where it is most evident. According to the ISH, the gesture-speech integration applies to all aspects of language comprehension, i.e. in addition to semantics, it occurs in the processing of pragmatic, phonetic as well as syntactic infor-

mation. However, the degree to which it affects the comprehension process varies. Building upon the body of evidence from gesture perception/comprehension studies, Kelly (2017) shows that it is (concrete) semantics, prosody and pragmatics where the multimodal co-expressiveness matters the most. In the comprehension of syntactic as well as (segmental) phonetic information and abstract semantic decoding, the role of gesture appears to be to a varying extent limited.

The research into the comprehension of the semantics of speech accompanied by gestures has mostly been oriented towards the effects of semantic discrepancy between gesture and speech on comprehension. If gestural and linguistic information is integrated into a semantic unit, conflicting or inconsistent "meanings" provided by the respective modalities should cause comprehension difficulties in addressees. Psycholinguistic methods developed for "unimodal" comprehension studies may be used here for behavioural measurements (typically reaction times in lexical decision or grammaticality judgements). Examples of particular research designs are given in Section 6.1. Kelly summarizes the generalizations that can be made based on the research findings of those studies, showing that the presence of iconic gestures enhances comprehension when the semantic information communicated by gestures is concrete, related to the physical or motor properties of concrete objects or actions (with gestures providing an additional focus on the object or event representation). When it comes to the more abstract, "disembodied" semantic information, conclusions about the role of gesture are not supported by a sufficient body of evidence – leaving the door open for further debate on the extent to which comprehension relies on embodied cognition-related processes such as mental simulation.

Strong evidence for an integrated comprehension system comes from the study of beat gestures in relation to prosody. In the Section 2.2, I gave an overview of the research of the role of gesture in prosody marking, providing a converging evidence of (i) a systematic temporal alignment between gesture and pitch accent and (ii) a systematic temporal shift between the peak of the gesture stroke phase and the frequency peak within the intermediate phrase.

In one of the first attempts to provide an empirical psycholinguistic base for the observation of gesture preceding pitch accent, Leonard and Cummins (2011) showed that when perceiving videos of co-speech gestures (beats) with the sound channel artificially shifted, the test subjects were more sensitive to the shift when the speech was manipulated to come before the onset of the gesture.

In his review of experimental evidence of gesture-speech integration at the prosodic level, Kelly mentions, among others, a study (Krahmer and Swerts, 2007) showing that beat gestures are not linked to prosodic emphasis in production, and also that when primed by input with beats, subjects perceive the same stimuli without beats as louder than stimuli previously unaccompanied by beats. Similar effects were also attested in phonetic processing at the segmental level. Bosker and Peeters (2020) report a case of what they call the "manual McGurk effect". Like in the original McGurk

effect (McGurk and MacDonald, 1976), they found a crossmodal influence of visual input on auditory perception at the level of syllables. Participants of their experiment perceived prosodic stress at syllables coinciding with beat gestures, which may under specific conditions also have an impact on the vowel quantity perception (in Dutch).

Closely linked to prosody is the gesture-speech integration related to the pragmatic aspects of information. Here, more than anywhere else, the role of gesture is evident: it is the gesture that often adds the critical piece of information attributing various speech acts to the same linguistic content. Kelly stresses the importance of (especially deictic) gestures in children's input in the acquisition of social-cognitive aspects of communication.

Syntax does not occupy a central place in the ISH. Kelly argues that the relative lack of gestural effects on syntactic processing indicates that gesture may be crucial only for some aspects of language comprehension. Syntactic and segmental phonetic processing could work independently of the embodied integrated system: "the automatized quality of phonemes and the conventionalized nature of syntax make processing (familiar) phonemic units and syntactic structures fast and easy" (Kelly, 2017, p. 260).

Being a general explanatory framework for the gesture's role in language comprehension, rather than a psycholinguistic model as such, the ISH takes into account the entire spectrum of gesture functions, i.e. it is not limited only to the "representational" gestures. According to this view, multimodal integration occurs obligatorily (Kelly et al., 2010, p. 266) – but only when the comprehension process can actually utilize the embodied aspects of language. When it does not, comprehension processes are claimed to hold onto body-independent cognitive processes.

Needless to say, evidence from speakers of a greater variety of languages is needed here. For instance, in the syntactic domain, gestural focusing may come into play more prominently in languages with freer word order than English. Kelly himself mentions a study suggesting that gesture facilitates parsing of ambiguous sentences in German (Holle et al., 2012).

3.2.3 Multimodal Construction Grammar

Gesture is inseparable from the production and comprehension of spoken language acros all stages of language processing and – at least to a certain extent – with respect to all structural levels of language. It is only logical, then, not to ignore the gestural component in the grammatical description either.

This "multimodal turn" in linguistics can be seen as a natural consequence of the paradigm shifts language sciences have been witnessing since the second half of the 20th century: starting with re-focusing on performance which marked the pragmatic turn (in all its facets across disciplines) and later leading to a series of partial perspective shiftings: e. g. from written to spoken language (Linell, 2005), from the view of language as a substantially homogenous phenomenon to the emphasis on linguistic di-

versity (Dahl, 1990; Evans and Levinson, 2009), from the introspective or "armchair" approach to truly empirical linguistics (Janda, 2013a), from language as a matter of individual competence to language as an intersubjective phenomenon (Geeraerts, 2016) and, finally, re-focusing from lexical units onto the idiomatic – or constructional – level (Fillmore et al., 1988).

As stated in the beginning of Section 3.2, the principles of CL, particularly its focus on language usage and the idea of embodied cognition, provide a theoretical ground for the inclusion of gestural expression into the grammatical framework. But, in fact, all of the paradigmatic shifts mentioned above constitute the general affordance of CL for "going multimodal". It especially applies to *Construction Grammar* (CxG), a grammatical framework falling under the CL umbrella, which resulted from a reflection of most of the above paradigm shifts, and which has in recent years branched out into the most elaborate of multimodal grammars.

Based on different grammatical frameworks, several attempts aiming explicitly at "multimodal grammar" have been made before. In his dissertation, Kasper Kok (2016) gives a detailed overview listing several functional and CL-inspired approaches to integrating gesture into the grammatical framework, while sketching his own model based on Functional Discourse Grammar, following the previous work of Connolly (2010). As both functional and cognitive perspectives converge in CxG and none of the other approaches has found a systematic application with a comparable reach, I will not recapitulate the advantages and disadvantages of various other models of multimodal grammar and will only focus on multimodal CxG here.

But before proceeding, I would like to address a point of contention which has been ignored by various functional and cognitive grammatical frameworks. *Role and Reference Grammar* (RRG, Van Valin, 2005) shares many basic assumptions with CxG (reviewed below) and as such is open for the integration of additional components, starting with prosody. But the main reason why RRG calls for the integration of not only prosodic, but also gestural information, is its formalism which, unlike CxG, retains the linear representation of syntactic structures of classical grammars onto which semantic and pragmatic layers are projected. The layered representation of RRG has found its application, i.a. in the description of topic-focus structures in crosslinguistic studies of information structure. In this context, a model of prosody representation in the RRG formalism was put forth by O'Connor (2008), providing an elegant and relatively fine-grained tool for a more adequate analysis of the profoundly multimodal phenomenon that information structure marking is. Despite being timely, O'Connor's contribution did not attract wider attention, nor has RRG witnessed any attempt to include gestural projection (which, in the context of information structure representation, would only be a logical next step).

Originating in the 1980s works of Charles Fillmore (e.g., 1988), CxG formed as a branch of CL in the early 1990s and currently exists in several versions,[70] all of which

[70]Version of CxG include Sign-Based CxG (Boas and Sag, 2012), Radical CxG (Croft, 2001), Berkeley

share three basic assumptions summed up by Fried and Östman (2004, p. 12):

> "(i) speakers rely on a relatively complex meaning-form patterns – constructions – for building linguistic expressions; (ii) linguistic expressions reflect the effects of interaction between constructions and the linguistic material, such as words, which occur in them; and (iii) constructions are organized into networks of overlapping patterns related through share properties."

With respect to other major grammatical theories, the notion of grammatical construction is innovative in abandoning the word level as the primary unit of reference, but it is also traditional in its definition as a linguistic sign in the Saussurean sense. But unlike the structuralist view of a sign as a rather rigid structure,[71] it should not be understood as a fixed combination of a certain form with a certain meaning, but rather as an organizing principle of the human conceptual system. Constructions are considered adequate analytic units for both linguistic and conceptual systems. CxG assumes that grammatical constructions should in principle serve as hypotheses for cognitive science to test how human cognition processes and stores linguistic representations. From the cognitive point of view, constructions are gestalts, which is manifested in one of their main features: non-compositionality. The gestalt principle that underlies the organization of human conceptual systems is based on the construal of "single complex object[s] from seemingly fragmented perceptual sensations" (Croft and Cruse, 2004, p. 63). As such gestalts, grammatical constructions can be identified at every level of complexity as non-compositional sets of formal and functional/semantic features. One of the revolutionary features of CxG is related to what counts as a feature constituting a grammatical construction. Unlike GG or the structuralist approaches, the CxG descriptive apparatus is not constrained by limited sets of predefined universal features (which also means that there is no place for empty categories), but the potential features are always bound to the context of usage. Being principally inductive (Fried, 2015), CxG draws on instances of actual usage, which allows us to account for a rich variety of (morpho-)syntactic, semantic but also phonetic and pragmatic features that could be of relevance in a particular instance. Although the list of potential features is open-ended in principle, the aim of CxG is to come up with (adequate enough) generalizations, which are customarily represented by box diagrams containing conventional annotations of *characteristic* features.

Strictly speaking, CxG has always been multimodal. By definition, the CxG formalism should accommodate any relevant information regardless of modality. To give

CxG (Fillmore, 2013) or Fluid CxG (Steels, 2017). Van Trijp (2013, p. 93) provides an overview diagram situating different versions of CxG within the development of major linguistics theories in the second half of the 20th century.

[71] Note that the label *structuralist* here refers neither to de Saussure nor Peirce, who stressed the dynamic and processual nature of a sign, but to the general focus on *langue* while paying no attention to social and psychological aspects of language that dominated continental structural linguistics.

an example, the inclusion of phonetic features (typically suprasegmental) is one of the distinctive trademarks of CxG. But, as we have already seen several times, prosody is just the beginning. For the sake of illustration, consider a construction in which the gestural component represents a necessary feature:

(7) I caught *a fish* ***this big***.

Similar to example (1) from the first chapter, this construction represents a productive pattern and seems to belong to a family of constructions that can be generalized as [*N-this-Adj*$_{size}$]. As gesture is not only a characteristic but also a necessary feature of this construction, it should be embedded in the CxG formalism. For the sake of simplicity, let us consider the construction *fish this big*, represented in a form of a box diagram[72] in Figure 14. The selection of features captured in the diagrammatic representation is only tentative: CxG operates with an open inventory of features – syntactic (*syn*), semantic (*sem*), pragmatic (*prag*), discursive (*disc*) or phonological – and the selection of the representative set of features needs to result of an *inductive* data analysis process.

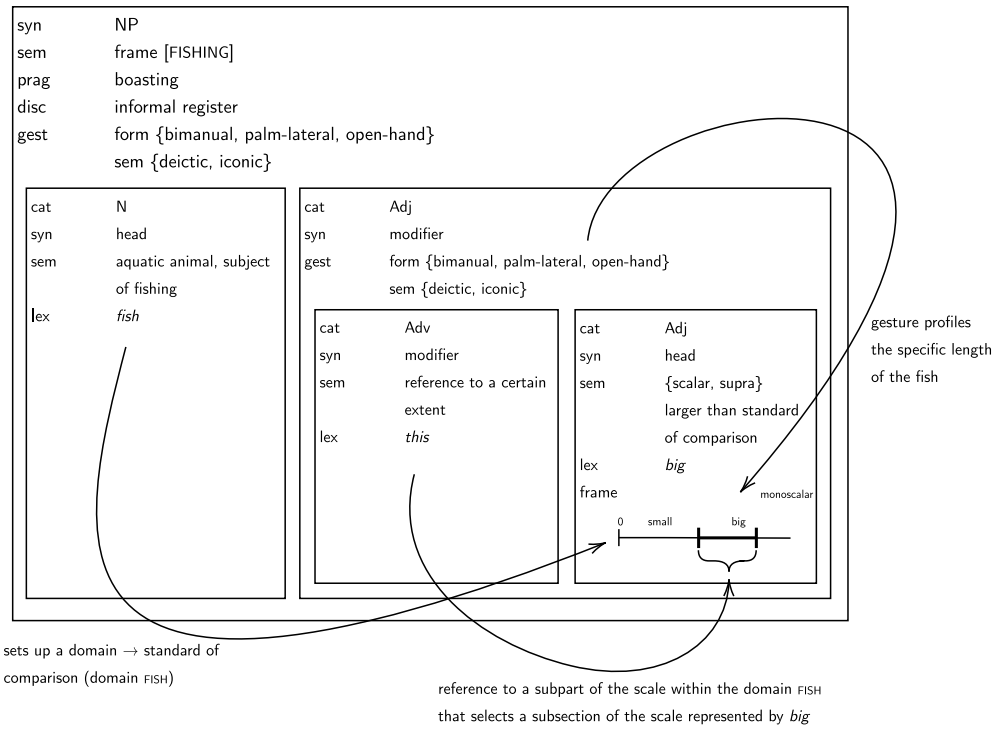


Figure 14: *Simplified diagrammatic representation of the* [fish this big] *construction*

In the proposed gesture-enriched CxG formalism, gestural features are represented under the label *gest* and comprise both formal (*form*) and semantic (*sem*) information, described via the defining phonological features (Bressem, 2013) and semiotic-functional dimensions introduced in Section 2.1, respectively.

[72] Overview of the conventions of the CxG formalism was provided by Fried and Östman, 2004.

Apart from boxes within boxes (each box representing a construction of its own) comprising of the defining features, the diagram is also complemented with arrows that highlight the dynamic intra-constructional relations: here, the arrows capture how the individual components of the constructions contribute to the scalar construal of the adjective *big*. *Fish* invokes a section of the domain FISH: the construction is used in the context of recreational fishing, prototypically with a rod and hook, which entails catch spanning a scale of fish sizes between a tiny tiddler and, e.g. a human-sized catfish. A relevant section of the scale (one of the frames of the domain FISH) is profiled by the adjective *big*, which is further specified by the adverbial use of the demonstrative *this*, selecting a subsection of the part profiled part of the scale. The subsection referred to by *this* is profiled by gesture, which iconically represents the specific size of the fish (or a reasonable exaggeration thereof). The two hands in the bi-manual palm-lateral open hand gesture represent the two extreme points of the fish's longitudinal axis, which corresponds to *object-centred depiction* (see Table 1). Together with *this*, the gesture co-refers to the profiled part the scale and as such, the gesture may be also described as *deictic*. According to Hassemer's and McLeary's (2018) typology of pointing gestures addressed in Chapter 2, this is a case of *surface edge demarcating*. Hence, the semantic properties of the gesture captured in the diagrammatic representation include both *deictic* and *iconic* values.

The first researcher to include gesture as an integral part of the construction (in the CxG sense) explicitly was Mats Andrén in his dissertation (2010), in which he coined the term *multimodal construction*. Andrén's work was focused on L1 acquisition, showing that some of the first constructions to be acquired by children, typically the so-called item-based constructions (Tomasello, 2005), are used multimodally from the very beginning. These constructions are accompanied by a specific gestural form in a non-compositional manner, e.g. a head shake with the lexical item *gone*.

Focusing on adult language, Steven Schoonjans, under the rubric of multimodal CxG, analysed non-manual gestures co-occurring with modal particles in German (e.g., *denn*, *doch*, *eben* or *einfach*) in a corpus of TV recordings. His analysis revealed that German modal particle constructions involve gesture with a varying relative frequency and that the degree of entrenchment of multimodal constructions depends on, among other things, the particle's specificity and grammaticalization.

Elizabeth Zima (2014) focused on recurrent patterns in the form of co-speech gestures accompanying motion constructions in English. Her analysis revealed that in the sample of nearly 400 instances of 4 different motion constructions ([V_{motion} *in circles*], [*N spin around*], [*zigzag*], [*all the way from X* PREP *Y*]), 60–80% included a conventionalized gesture form. Her data were also extracted from TV recordings.

According to Zima, when a co-speech gesture exhibits a certain degree of recurrence, conventionality, and non-compositionality, it might be considered a part of the given grammatical construction. The question that necessarily follows is how one can establish benchmarks for these parameters. Schoonjans (2017) addressed the theoret-

ical and methodological problems of multimodal CxG, identifying four major issues:

(i) how to establish a recurrent gesture – construction pattern given that gesture is a "fluid" phenomenon and its particular instances share only a partial set of features;

(ii) how to set a frequency threshold for a multimodal construction to be considered entrenched or conventionalized;

(iii) how to deal with the lack of alignment between the gestural and verbal components of multimodal constructions;

(iv) how to tell that a multimodal construction is a real cognitive unit, i.e. whether it is processed as multimodal.

Resolution to these issues still represents a challenge for the future development of multimodal CxG. Nevertheless, Schoonjans' point is that each of these issues applies to unimodal CxG with the same seriousness as well, and so "these issues should not be used as arguments against including multimodality in CxG, but rather as incentives to rethink and refine ideas about CxG in general" (p. 6).

At a more practical level, the central problem related especially to issues (i) and (ii) are *data*. Unimodal CxG hugely relies on corpus data, but multimodal corpora are still no match to referential corpora of written language, nor are comparable to representative spoken corpora, such as the spoken component of BNC[73] or the Czech ORAL corpus.[74] While there is still a way to go for multimodal corpora in general, workable options can be found. One of them is the *UCLA Library Broadcast NewsScape*,[75] also referred to as the Red Hen Archive,[76] a repository of television broadcast in English and a number of other languages (Turner and Steen, 2013). Crucially, the Red Hen Archive is ever-growing (as of now, it consists of approximatelly five hundred thousand[77] hours of video) and although it is not strictly speaking a language corpus, it is supplemented with texts (closed captions) and is searchable via a web interface. Thanks to that, the Red Hen Archive has become a data source for several of the current undertakings in multimodal CxG. One important aspect of the Red Hen data is its ecological validity – TV recordings are a collection of various types of data, ranging from very formal to a relatively spontaneous production. This aspect will be addressed in Section 5.1. Another aspect is that the Red Hen Archive allows for large dataset extraction (at least theoretically, depending on the research question), which is crucial for any quantitative work within CxG.

[73] http://cass.lancs.ac.uk/cass-projects/spoken-bnc2014/

[74] https://www.korpus.cz/

[75] http://newsscape.library.ucla.edu/

[76] Called after *The International Distributed Little Red Hen Lab* (https://sites.google.com/site/distributedlittleredhen/home) – a consortium of researchers who collaborate on the research tool development and carry out analyses of the data from the archive.

[77] As of 2020.

So far, not many studies conducted under the rubric of multimodal CxG have ventured beyond qualitative accounts or reporting relative frequencies. Recently, harbingers of the future multimodal CxG research have appeared, using big data extracted from the Red Hen Archive. To give an example, Pagán Cánovas et al. (2020) or Uhrig (2019) applied collostructional analysis (Stefanowitsch and Gries, 2003) to measure the cross-modal association in multimodal construction candidates – a method that requires a relatively high number of observations.

3.3 Bridging the gap: a unified account

Recently, a number of attempts to reconcile the two seemingly incompatible approaches to embodiment has emerged (discussed above). On the one hand, the very framework of CA faces a kind of methodological dead-end unless it opens itself up to quantitative methods. The absence of the quantitative aspect has been a significant burden for traditional CA in terms of wider reproducibility of its analyses and their usefulness in cross-disciplinary perspective, given that these two aspects are considered a *sine-qua-non* across social sciences. With the adoption of precise quantitative procedures, based on hypothesis testing using complex statistical modelling, and taking the state of the art in related disciplines into account, the interactionist research indeed produces outcomes appealing to social scientists beyond the CA community.

On the other hand, there is not much of the *mental states* left for the current cognitivist either. The more the "window to the mind" opens due to neurological findings, the less one has to recourse to mentalist constructs: the range of cognitive processes and phenomena that we are able to reduce to direct bodily reactions (mostly at the neural level) is widening. To recall what was stated above, in this regard the theories of embodied (social) cognition succeed in facing the problems expressed by the behaviourists, problems that were not adequately resolved by the first generation of cognitive scientists. Such approach to embodied cognition is represented, e. g. in the works of Shaun Gallagher (Gallagher and Hutto, 2008) or – as for the CL-based linguistics – scholars like Arie Verhagen (2005; 2015).

In this section, I will briefly review two recent contributions to this debate: one by experimental psychologists (de Ruiter and Albert, 2017), the other by a gesture researcher originally representing the interactionist approach (Streeck, 2015). The debate has been open for a while, drawing attention, among other issues, to the re-evaluation of the status of mental representation in cognitive science (Clowes and Mendonça, 2015), to re-introducing phenomenological perspective to cognitive science and to CL in particular (Zlatev, 2016) or to the convergence of construction grammar (part of the CL-theory complex) and CA (Fischer, 2015).

Psychologist J.P. de Ruiter and computer scientist and psychologist Saul Albert (2017; 2018) offer a view on this schism taking a methodological perspective. Although

the field in focus is experimental psychology, the argumentation is applicable to the cognitivist approach to gesture studies too, for it is also considered a part of cognitive science with its empirical commitment (that is usually realized in the form of experiments).

The authors emphasize that both CA and experimental psychology share the empirical stance rejecting introspection as a source of data. The division line between the two approaches lies in the general theoretical-methodological paradigm they adopt. The authors argue that whereas experimental psychology paradigms are committed to Popper's falsificationism (Popper, 1935), CA relies on a completely different assumption. The notion of falsification according to Karl Popper refers to the quality that a theory must possess in order to be considered scientific. If a theory allows for being falsified on the basis of any kind of empirical data, i.e. it can be subjected to a test (e.g. an experiment), it is justified to be maintained until its eventual debunking.

Rather than fitting theories into a framework of hypothesis testing based on a set of area-specific methods, CA "avoids invoking entities of concepts that are not firmly grounded in natural observation" (de Ruiter and Albert, 2017, p. 96). The only acceptable source of data for CA is the observation of naturally occurring interactions, and interpretation is limited only to instances of observation of an interaction. Thus, an observation cannot lead to a context-independent generalization that would contribute to a formulation of a theory of human interaction. Where there cannot be a generalized conclusion drawn from an observation, there is also no use for quantification (Garfinkel, 1967), making CA an exclusively qualitative methodology.

The authors find one of the prospects for convergence of the two methodologically disjointed approaches in connection to the so-called replication crisis that arose originally in psychology in the early 2010s but has spread into social sciences (Pashler and Wagenmakers, 2012). The fact that a majority of published empirical studies in social sciences fails to be successfully replicated under comparable conditions may be related to "the wide conceptual gap between the *context of discovery* and the *context of* [its] *justification*" (de Ruiter and Albert, 2017, p. 97, emphasis JJ) that opens during transfiguring a real-world observation into a set of quantifiable variables in order to test it. In the study of language interaction, this gap is manifested when the common practices of empirical (experimental) research are applied to interactional data. The common experimental procedures are the context of justification, which is bound by assumptions (e.g. those about the notions related to cognition) that might be relied upon in experimental psychology as if they were based on attested findings but that, in fact, originally come from someone's introspection. The context of discovery involves the nature of interactional data characterized by a plethora of multifaceted and (in principle) non-reproducible phenomena (as it involves "the unobservable": the actors' intentions and construals). The context of justification should be grounded in the context of discovery, although in practice it is often bound in pre-existing theories and models "without reference to the everyday interactional situations that presum-

ably give rise to the psychological effects and mechanisms being studied" (Albert and De Ruiter, 2018, p. 2). The authors offer a set of remedies applicable across the entire process of interactional data analysis. In general, researchers should strive to diminish the discrepancy between the context of discovery and the context of justification by improving the ecological validity of their coding. This can be achieved by the introduction of the "usage-based", formal coding units constrained by a prior qualitative inspection of the material. Researchers should draw upon the practices and standards developed by CA to capture as many relevant aspects of the interaction as possible. Importantly, as the authors note, the CA transcription technique (see above) was developed to reflect the "resources available to the participants" in the context of interaction. The final data analysis should then be subjected to repeated informal peer review that, on top of the apparent benefits, represents another case of interactional setting which provides a floor to check the assumptions in a conveniently practical fashion. CA-oriented researchers, on the other hand, can and should benefit from experimental psychology as well. The authors point to the issue of multimodal analysis of interactions as an example of where experimental approaches might neatly complement ethnomethodology:

> "We argue that experimental methods for predicting and testing generalizations could offer a […] productive experimental counterpart to the challenges of analysing embodiment, which would benefit CA by highlighting findings that have been, for whatever reason, inaccurate of inadequately specified" (de Ruiter and Albert, 2017, p. 99).

As examples of how these principles can be successfully implemented, de Ruiter and Albert list a number of recent studies (Holler et al., 2016; Kendrick and Torreira, 2015; Dingemanse and Enfield, 2015; Dingemanse et al., 2015). Most of these studies were focused on the flag-ship topic of CA: turn-taking. However, the novel approach departs from the traditional CA accounts of turn sequence organization in conversation (starting with Sacks et al., 1974, followed by countless case studies) in several aspects. It is based on (multimodal) corpora of naturally occurring conversations (spoken as well as signed) that were not only transcribed but were also coded in annotation software allowing for time measurements with millisecond precision, leading to diverse sets of data suitable for complex statistical modelling and, at the same time, characterized by high ecological validity. Moreover, it has been shown that the CA-framework can be successfully adopted in studies using eye-tracking (Holler and Kendrick, 2015) or even neuroimaging (Bögels and Levinson, 2017).

The present study will draw upon the above principles, even though they are primarily directed at confirmatory research and this study is exploratory. In the discussion (Chapter 7), I will elaborate on other procedures (applied in the present study) leading to improving ecological validity throughout the data analysis process.

Similarly, although not primarily concerned with methodology and dealing with a more specific topic, Jürgen Streeck (2015) deals with the issue of convergence between cognitive and interactionist perspectives on embodiment. Importantly, it has to be made clear that for Streeck, the two parts of the dispute are not framed in the same manner as they are for de Ruiter and Albert.

Streeck offers a view of integration between the interactionist paradigms and "the new conception of the living body emerging in biology, cognitive- and neuroscience, and sociology and anthropology" (Streeck, 2015, p. 420). This "new conception" is the one that draws on the idea of embodiment in a direct, anti-dualist sense.

Finding a common language is not the only crucial matter for such integration. Whereas it is necessary to find a shared methodological ground in order to bring CA and experimental methods closer together, there is already an agreement between the two camps upon the need for abandoning the dualist view of cognition. However, as Streeck points out, the dualist heritage pervades beyond that. Even though communication is approached as embodied in the sense that the communicating agents are bound by their bodily experience which drives the transmission of meaning, the very focus on the "subject" as a central concept in communication is burdened by the dualist perspective, while the "notion of the body as passive bearer of coded cultural messages has […] been a tacit foundation of much work" (*ibid.*). The alternative paradigm is based on the concept of *intercorporeality* introduced by Merleau-Ponty (*intercorporéité*, 1962):

> "As the active body acquires skills, those skills are stored, not as representations in the mind, but as dispositions to respond to the solicitations of the situation. The living body is constituted not only by its incorporation of things, but also by its incorporation of other bodies" (Streeck, 2015, p. 422).

In the view proposed by Streeck, intercorporeality is thus replaced by the conception of *intersubjectivity* – "sharing of aective, perceptual and reective experiences between two or more subjects" (Zlatev, 2008, p. 215) – replacing subjects with living bodies. Streeck gives two examples of theoretical frameworks for the study of communication and its multimodal aspects that comply with the idea of intercorporeality. The first one is Edwin Hutchins' theory of distributed cognition (1995; 2006), a holistic view of cognition as a complex network (*cognitive ecology*) constituted by interaction and organization of human agents, modality of their communication, tools they handle and space they occupy and other contexts that take part in providing meaning to the communication (one may recall Lotman's notion of *semiosphere* (1984)). Among the so-called *cognitive artifacts* (i.e. particular aspects of cognitive ecology), gesture is salient as it serves not only for manipulation and handling but also for conceptual representation. As Streeck, who reflects this dual nature of gesture in his gestural ecologies (see above), emphasizes, "instrumental and communicative action overlap and do not constitute separate 'modules'" (Streeck, 2015, p. 430).

The other framework is represented by Charles Goodwin and his approach to CA. Streeck notices a shift in Goodwin's work from the traditional ethnomethodological focus on subjects in interaction to "collaborative activity [defined] as [a] primordial site of sociality" (*ibid*., p. 431). In Goodwin's recent work (e.g. 2013), we are presented with an analysis of action organized by collaborating subjects who are equipped with a certain *contextual configuration*, i.e. a set of semiotic resources (e.g. "language structure, categories, prosody, postural configurations, the embodied displays of a hearer, tools, etc." (Goodwin, 2013, p. 21), and who are situated within *co-operative transformation zones* – culture-specific patterns of action organization in which the appropriate resources are combined. By repetition, co-operative transformation zones are reinforced, but they are also constantly transformed. "The pervasiveness of co-operative transformation zones constituted within the midst of ongoing action is central to the accumulative organization of human culture, knowledge and social life" (*ibid.*).

Again, we see an attempt to "disperse" the subject into the environment, stressing its situatedness in context and its dependence on the communicative tools given in the locus it "dwells" in (in the Heideggerian sense) instead of relying on the subject as the primary means of explanation. Be it from the perspective of embodied cognition or embodied action, the goal of such embedding of the subject is the same: to come up with a model of communication that would be more appropriate than the traditional conception of the speaker as a "central processing unit" (2015, p. 240) that makes use of non-verbal means of communication such as manual gestures as mere secondary tools. Nevertheless, as I noted above regarding Merleau-Ponty's conflated metaphorical frames, the dualist view is hard to overcome for it is deeply rooted not only in folk psychology but also in the way cognition is conceptualized in scientific discourse. As Streeck concludes, the

> "real difficulty at present appears to be finding a postdualist language to formulate our understanding of the communicating human body and of the ways in which human bodies understand one another in social interaction" (*ibid*., p. 433).

In the vein of the methodological convergence sketched in this section, an interesting development is under way within the CxG framework. For some time now the CxG proponents have pointed to the general compatibility of constructional approaches with CA (Fried and Östman, 2005) or even suggested practical solutions for combining the two methodologies (Linell, 2009; Fischer, 2010, 2015; Nir et al., 2014). Some of these contributions have been carried out under the flag of *Interactional CxG* (Wide, 2009; Imo, 2015; Hsieh and I-Wen Su, 2019). Alternatively, a label *Dialogical CxG* was proposed by Brône and Zima (2014). Enriching constructions with interactional aspects can basically take two (complementary) forms:

(i) expanding the pragmatic component within constructions at the clause level by attributes related to interaction, as well as specifying the morphosyntactic and phonological components with respect to the particular discursive function

(ii) expanding the scope of constructional description to multiple conversational turns

The approach (i) has mostly been adopted with respect to the potential of a particular construction to enter sequential discourse relations. The constructions in focus involve phenomena such as anaphora, various discourse markers or cases that Linell (2009) calls *responsive constructions*. The second perspective (ii) is represented, e.g. by the studies of the *discourse resonance* phenomena (Du Bois, 2014). Resonance (at various structural levels) emerges in juxtaposition to conversational turns when a certain pattern echoes across several utterances by several participants, giving birth to "new, higher-order linguistic structure(s). Within [these] structure[s], the coupled components recontextualize each other, generating new affordances for meaning" (*ibid.*, p. 360).

In the light of the above, it is apparent that the inclusion of gesture is crucial for both Interactional CxG approaches. Steps towards the reflection of gestures from perspective (i) can be traced in studies of the referential so in German (Ningelgen and Auer, 2017) or German modal particles (Schoonjans, 2014).

The dialogical approach (ii) can benefit from considering how discourse resonance is marked by gesture. For instance, gestural resonance may be signalled by the presence of the so-called *catchments*, discourse-cohesion devices "recognized from recurrences of gesture form features over a stretch of discourse" (McNeill et al., 2015).

*
* *

Based on the integrative approach outlined above, a general assumption concerning the very definition of gesture will be made at this point. Following the metaphorical frame introduced in the opening of the present chapter, gesture will henceforth be understood in terms of the "Copenhagen interpretation". In other words, gesture will be approached as an entity of dual nature – on the one hand, it is a continuum, inseparable from its immediate context; on the other hand, it is natural to frame it as a "countable" unit. The latter is nothing but an *approximation*. This approximation must principally be made on the basis of a quantitative assessment.

Measurement of the degree of agreement among annotators who code gestural units is one of the ways this assessment can be carried out. Kita, van Gijn and van der Hults (1998) showed that humans are capable of arriving at very similar units when independently coding gestural movements, given a set of predefined features. Inter-individually, an agreement can be reached about (approximate) division lines between gestural units. In fact, this is in line with one of the basic assumptions of cognitive linguistics (inspired by Gestalt psychology (Koffka, 1935)) about the nature of our percep-

tion which often translates continual natural phenomena into quanta – *gestalts* upon which human perception works.

At the same time, what is missing from the approximation must be stressed too, constantly bearing the observer paradox in mind (and the fact that it cannot be avoided, at least not with the analytic apparatus that we are bound to deal with at the moment).

4. Gesture and eventuality

> *"[S]elva oscura sometimes known as aspectology…[is full of] obstacles, pitfalls, and mazes which have trapped most of those who have ventured into this much explored but poorly mapped territory…"* (Macaulay, 1978, p. 416–147).

It is not only the domain of aspect and aspectuality that often gives the impression of a dangerous and unfathomable area: the same is true about the entire domain of event semantics where the troubled subject of aspect resides. Not to go astray right in the beginning, let me first clarify the usage of terminology related to this domain and, by doing so, to introduce the concepts that are central to this study.

The following chapter gets to the heart of the matter, which is a crosslinguistic study of co-speech gestures in relation to the linguistic expression of eventuality. It prepares the ground for the empirical part of this study (Chapters 5 and 6):

First (Section 4.1.1), I will make a brief note on terminology, defining some basic concepts related to the area of events semantics, as they are handled within this study. Then I will review the current state of the art in the research of co-speech gestures with respect to eventuality, beginning with a relatively well-studied semantic domain of motion (4.2.1). The area of interest of this study – the research of co-speech gestures and (broadly defined) aspectuality – is reviewed in Section 4.2.2 followed by an overview of the empirical evidence from the psycholinguistic research on the expression of eventuality in sign languages (4.2.3). Finally, the cognitive approach to event semantics will be introduced (4.3.1), that allows for an analysis of the *multimodal construal* of events based on the concept of *aspectual contour*. The research questions and assumptions of the present study are outlined in Section 4.3.3.

4.1 Event semantics

What does the titular notion of *eventuality* actually stand for and where did it come from? Using this term to label the area that is covered in this work is not a common practice – but I resort to it as there is no term at hand that would be suitable as well as agreed upon. The area of interest is referred to by a number of terms, often used as the *pars pro toto* labels for a more general category besides its more specific meaning: *event semantics*, *situation types*, *aspectuality*, *eventualities*, *Aktionsart* or *predicate semantics*. All these labels refer to the domain of the verb semantics, although each from a different perspective and with a different scope. I will briefly review the pitfalls hidden behind them. The following section will provide only a very rough review of terminological systems one can encounter in the domain of verb semantics. I do not attempt to give a comprehensive account of the theoretical background of verb semantics from a

linguistic-general perspective here, as it is beyond the scope of this study.

4.1.1 Typology of events

In the domain of verb semantics, ***events*** comprise a broad range of categories of verb meaning, not only "events" in the narrow sense (i.e., evoking a punctual, bounded or "one-off" phenomenon) but actions, processes, states and whatever else may fall under the umbrella of verb semantic categories.[78] These categories are a result of dividing up the semantic domain of verbs based on the character of events' unfolding in time.

A number of existing classifications that do comprise events as particular category (e.g. Bach, 1986; Daneš, 1971; Smith, 1997) and therefore it is not very fortunate to use this terms as a hyperonym.

In the cognitive accounts, the term *event* is often used as it is understood in cognitive psychology, i.e. as a fundamental unit of human experience. Perception, memory and conceptualization is organized in events: episodic segments of the stream of consciousness. According to the Event Segmentation Theory (Kurby and Zacks, 2008), events are stored representations of the immediately happening experience, based on probabilistic learning of prediction accuracy (so-called *event model*). The event boundaries are the points where people "update memory representations of 'what is happening now'. The processing cascade of detecting a transient increase in error and updating memory is perceived as the subjective experience that a new event has begun" (Kurby and Zacks, 2008, p. 72). The perception of the event boundaries (i.e. the onset of the process of updating the event model in working memory) has its distinctive neurological signature marked by activation in the visual cortex, posterior medial cortex and *gyrus angularis* (Baldassano et al., 2017).

In this study, events are assumed to be subject of *construal* (see 2.2), which is not in opposition to the Event Segmentation Theory. In fact, the multimodal construal approach builds upon the basic general-cognitive operations and event segmentation should be counted as one of those operations, perhaps even as a cognitive process that underlies all the four kinds of the construal operations, as all involve further profiling of the segmented events. In the multimodal construal view, boundaries of events may be signalled by gesture. Event models may involve stored embodied (motoric) schemata (which is supported by experimental evidence by Zacks et al. (2009), who found that changes in movement patterns are aligned with event segmentation). Crucially, the link between perception and language (or, to be more precise, what Slobin called *lexicalization patterns*, i.e. entrenched, language-specific constructional patterns (discussed in 4.1.2) is not uni-directional: language is not only representation of conceptual structures based on perceptual experiences, but, as will be illustrated below, it

[78] Maybe the most prominent is the classification by Zeno Vendler (1967) who divides up the English verbs into four categories based on what he calls time schemata (ways of an event's unfolding in time): *activities*, *states*, *achievements* and *accomplishments*.

may substantially affect how and what we perceive.

Smith (1997) uses the umbrella term ***situation*** (introduced by Comrie, 1976) and distinguishes the following situation types: states, activities, accomplishments, semelfactives and achievements, "according to [an event's or state's] temporal properties" (p. 3). Besides situation, Smith also recognizes viewpoints (perfective, imperfective, and neutral), thus mirroring the traditional distinction between aktionsart (or lexical (or semantic) aspect, defined by Comrie as lexicalised semantic distinction or "inherent" meaning (1976, pp. 6–7) and aspect (or grammatical aspect, defined as the grammaticalized "internal temporal constituency of [a] situation" (*ibid*., p. 5)).

Daneš (1971), in his classification of verb semantics, defines *situations* as static types (opposed to dynamic *actions*).

The term *situation* is also prominent in the cognitive approaches. For instance, the term *situation models* refers (in the context of discourse processing) to "the construction of a mental representation of the state of affairs denoted by that text rather than only a mental representation of the text itself" (Zwaan, 2016, p. 1028). *Situation types* may also denote the configurations of semantic roles at the level of argument structure (Fillmore, 2012). Croft (2012) understands *situation types* as a general term for "any semantic structure denoted by any utterance or semantically coherent part of an utterance" (p. 405), whereas *Aspectual types* represent types of *construals* of an *event* in terms of its *aspectual contour*: "the sequence of PHASES representing how a particular event is construed as unfolding over time" (see 4.3.1).

Predicates may consist not only of verbs but also of other parts of speech[79] (copular constructions, mostly associated with states). It might be therefore more suitable to take a syntactic perspective rather than lexico-semantic and talk about *predicate semantics* rather than verb semantics.

Two problems arise with predicate semantics, though. First, there is again a potential for terminological confusion as predicate semantics also refers to a quite different[80] domain, namely formal logic. Second problem is more serious. Eventuality encoding is not only limited to the predicate (VP and its arguments), but it can also be encoded in other NPs, AdjPs or AdvPs completely outside of predicate. This is the case of deverbatives that inherit semantic features of eventuality from the source verb (with varying degree of semantic change occurring in process – see e.g. Lehečková, 2011).

In Bach's (1986) classification of verb semantics, the superordinate category is ***eventualities***. Eventualities are divided up into states and non-states which further break up to processes and events. There are two types of events in Bach's classification: *protracted events* (e.g., *walk to Rome*) and *momentaneous events* (*happenings* such as *notice*

[79]Let alone the fact that the very concept of *part of speech* is problematic from the cross-linguistic perspective.

[80]But not completely remote either – the difference dwindles especially when it comes to formal semantics.

and *culminations* such as *die*). The term eventuality is, in this sense, primarily used in the formal accounts of verb semantics (Filip, 1999, *inter alia*).

Finally, the term ***Aspectuality*** serves as an umbrella category for both grammatical aspect and Aktionsart. In their crosslinguistic study (see Section 4.2.2, Alan Cienki and Olga Iriskhanova defined aspectuality as:

> "the cognitive ability to construe events in alternate ways – as having boundaries specified or not specified, or as being telic or atelic, or as being a complete whole or a process unfolding in time. This ability is verbally expressed by the category of aspect. There are two basic ways of conveying aspectuality in language – through grammatical aspect […] and/or lexical aspect […]" (Cienki and Iriskhanova, 2018, p. 179).

Such a term seems to be very convenient (and is widely used) as it captures a greater variety of semantic categories (regardless of the degree of grammaticalization), however, it is still limited by the scope of the term *aspect*, that is primarily associated with the temporal-perspective construals. But events may be construed in alternate ways also outside of the temporal domain: event structure also involves construals of locus – reflecting another cognitive ability: to construe events in various spatial or structural configurations (e.g. FIGURE VS. GROUND, (Langacker, 1987a; Talmy, 1972) or frames of reference (Levinson, 2003)). Finally, events are also construed in terms of MOTION (Talmy, 1985) and other force-dynamic construals based on image schemata such as CHANGE OF STATE OR TRANSFER (Croft et al., 2016). Therefore, in this text, the term ***eventuality*** will be used as a superordinate category subsuming the temporal, spatial and force-dynamic aspects of event structure.

4.1.2 Conceptualization of event structure

Crosslinguistic differences in the speakers' and signers' event-conceptualization strategies have represented one of the most addressed topics in the new wave of research inspired by the concept of *linguistic relativity*. Once refuted as pseudoscientific (Lenneberg, 1967; Pinker, 1994), the idea that knowing and using a specific language affects non-linguistic cognition, associated with Benjamin Lee Whorf (1941), found its way back to the mainstream scientific discourse in the 1990s (Lucy, 1992; Gumperz and Levinson, 1996). Since then, numerous psycholinguistic studies provided robust evidence of relativistic effects in various cognitive domains: temporal cognition (Boroditsky, 2001), spatial cognition (Levinson, 2003) or colour perception (Thierry et al., 2009). Dan Slobin focused on the differences in conceptualization of events involving motion. In particular, he focused on how language-specific constructions expressing simple motion events (such as *a boy climbs a tree*), lead the speakers of the given language to paying attention to different aspects of the event presented in a form of non-linguistic (e.g. visual) stimuli. Slobin research was based on Talmy's (1985) semantic typology concerned with the expression of motion, distinguishing between

the so-called Verb-framed and Satellite-framed languages.[81] In the Verb-framed languages (or V-languages, e.g. Spanish, Japanese, Korean or Turkish), the semantic component of PATH of motion is encoded in the verb. Languages that frame the motion's PATH not on the verb but outside the verb, on a so-called *satellite*, a syntactic category that comprises all kinds of constituents dependent on the verb,[82] are called Satellite-framed languages (or S-languages, such as English, German, Mandarin or Czech). See the examples of expression of the same motion event (*I rolled the keg into the storeroom)* in a V-language (8) and a S-language (9):

(8) Spanish (adapted from Talmy, 1985)

Met-í	*el*	*barril*	*a*	*la*	*bodega*	*rodándo-lo*
move-in-1SG.PST	ART	keg	to	ART	storeroom	roll.PTCP-it.ACC
(PATH)		(FIGURE)			(GROUND)	(MANNER)

'I rolled the keg into the storeroom.'

(9) Czech

Do-kul-il	*jsem*	*sud*	*do*	*sklep-a*
PRF-roll-1SG.PST	COP	keg	into	storeroom-GEN
(MANNER)		(FIGURE)	(PATH)	(GROUND)

'I rolled the keg into the storeroom.'

Slobin showed that the different encoding of motion events in S- and V-framed languages may lead their speakers to different interpretations of events presented non-linguistically. Speakers of English, German, Dutch and Russian (S-languages) and Spanish, French, Turkish and Hebrew (V-languages) (Slobin, 1996, 2004), when presented with the same pictures depicting motion events and asked to describe them, focused on different aspects of the movement. Unlike the speakers of S-languages, the speakers of V-languages did not specify the MANNER of movement in their narratives, focusing only on the movement's PATH, i.e. they used predominantly PATH-verbs like the French *sortir* ('exit') or Turkish *çıkmak* ('exit'). This was, according to Slobin, due to differences in speakers' "habitual attention to manner of motion" (2000, p. 113), linked to different patterns of how motion event structure is encoded in grammatical constructions.

[81] This is not to say that all languages should belong either among V- or S-framed groups. As the sample of languages analysed in the framework of Talmy's semantic typology expanded, atypical cases emerged and the original theory has undergone several revisions (Slobin, 2004, who introduced a third category: "Equipollently-framed language" or E-language, eg. Thai (Kam-Tai, Tai-Kadam, Thailand). Generally, the original distinction developed into a scalar view with S- and V-languages as the extreme poles of the continuum

[82] Or to be more precise, on the verb root, as the verb's affixes also fall under the category of satellites.

Slobin argues that the different interpretation cannot be explained by linguistic relativity in the Whorfian sense (i.e. a grammatical system influencing language-independent cognition) but it rather manifests different ways of what he calls *thinking for speaking* – "a special form of thought that is mobilized for communication" (Slobin, 1996, p. 76).

Slobin's findings inspired a series of comparative studies of event conceptualization, with a particular focus on *motion events*,[83] building upon various behavioural methods of examining the linguistic effects on *non-linguistic* cognition in mono- and bilingual speakers. The pool of languages put under scrutiny includes languages that differ in terms of MANNER/PATH framing (English vs. Greek, (Papafragou et al., 2008), Swedish vs. Spanish (Bylund, 2009)) or in terms of presence of a grammaticalized category of aspect (English vs. Swedish (Athanasopoulos and Bylund, 2013), English vs. Swedish and Afrikaans (Bylund et al., 2013), Slovak vs. Hungarian (Januška, 2017)). The study by von Stuterheim et al. (2012) included three clusters languages with different distribution of imperfective and progressive: Czech, Spanish, Standard Arabic and Russian vs. German and Dutch vs. English, see also Mertins, 2018, for a review). Visual stimuli (either animations or videoclips) capturing simple directed motion events were used to elicit verbal descriptions of the motion events and also in non-linguistic tasks (e.g. *similarity judgment task* or various memory tasks). Eye-tracking techniques were also employed in the studies by Papafragou et al. (2008) and von Stutterheim et al., (2012) to examine the (fixations of the) eye movements of the participants during both linguistic and non-linguistic tasks.

The evidence converges towards a general tendency of speakers to a) adopt different focusing strategies in the verbal descriptions of directed motion events according to grammaticalization patterns in the respective languages, b) to focus on different aspects of visually represented motion events in non-linguistic behavioural tasks, i.e., in case of S-/V-framing, the speakers of V-languages such as Greek first focus on PATH-related aspects of a scene, whereas speakers of S-languages such as English focus on MANNER-related parts of the scene first. However, this effect was observed only in the non-linguistic conditions involving the activation of the memory-related processes (Papafragou et al., 2008; Trueswell and Papafragou, 2010). In case of grammaticalized aspect, the non-linguistic evidence suggests a tendency of the speakers of aspectless languages to adopt what von Stutterheim et al. (2012) called a *holistic perspective* in which they focus on the endpoint of a motion event, even when it is not explicitly captured in the stimulus video. Speakers of languages with some kind of grammaticalized progressiveness or imperfectivity category, on the other hand, adopt the so-called *phasal perspective*, focusing on the event's progress.

Findings concerning motion event conceptualization in speakers of English and Czech are highly relevant in context of the present study: it was shown (von Stutterheim et al., 2012; Mertins, 2018) that whereas the English speakers have a tendency for

[83] Specifically, most of the studies focused on directed motion constructions ([V_{motion} PREP$_{dir}$ *LOCATION*], see De Knop (2020) for a constructionist account.

the phasal perspective as expected given the presence of grammaticalized progressive aspect (in the progressive verb forms), the Czech speakers do not exhibit the same tendency. In fact, the Czech subjects performed in a very similar manner to the German group, i.e. speakers of a language with an expected preference for a holistic perspective. The speakers of other Slavic languages in question (Russian, Polish and Slovak) did exhibit the tendency for phasal perspective. This striking difference was linked to a specific position of Czech aspectual system among the other Slavic languages which might be due to a long history of language contact between Czech and German. I will expand on this issue in Chapter 5.

4.2 Embodied eventuality

4.2.1 Gesture and motion events

The majority of the explorations in the area of eventuality with respect to gesture focuses on the relation of motion event framing and gesture from a crosslinguistic perspective. Initially, this direction of research was inspired by Slobin's notion of *thinking for speaking* introduced above. The question then arises as to whether the different (language-driven) interpretations of the same reality also comprise use of different gestural patterns is particularly important for the study of co-speech gestures' role in language production. If "thinking for speaking" is reflected in gesture then it is legitimate to assume that gesture and speech share some kind of a common production mechanism. The studies of "thinking and gesturing for speaking" focused on motion events have indeed provided the empirical grounding for psycholinguistic models of gesture-speech production interface (see Section 3.2.2).

A series of studies carried out by Aslı Özyürek and Sotaro Kita (Kita, 2003; Kita et al., 2007; Özyürek et al., 2005) represents a hallmark of the research into the the role of gesture in structuring the motion events. Özyürek and Kita compared co-speech gestures occurring with motion event expressions in English, Turkish, and Japanese. Turkish and Japanese are considered V-framed languages. The authors argued that co-speech gesture – as well as speech – are product of a spatial cognition–speech production interface and thus do not only encode (non-linguistic) spatio-motoric properties of the referent, but also structure the information about the referent in the way that is relatively compatible with linguistic encoding possibilities (Kita, 2003, p. 17).

According to this view, framed as the Interface Hypothesis (see 3.2.2), gestures in Japanese and Turkish were expected to encode MANNER and PATH of motion differently compared to English.

In their study (as reported in Kita, 2003), the method was again adapted from McNeill and Levy (1982) – narratives were elicited from 16 native speakers of American English, 18 Turkish and 17 Japanese native speakers. As in the above addressed McNeill's research, this study has also revealed similarities between gestural and lin-

guistic encoding of motion events. Kita and Özyürek observed the difference between the three languages in terms of the number of separate gestures used for the expression of MANNER and PATH. In this study, the Turkish and Japanese speakers produced predominantly separate gestures for MANNER and PATH whereas English speakers were more likely to produce a single gesture expressing both aspects of the motion event.

A similar pattern was reported also in comparison between speakers of Danish, a S-framed language, and Italian, a (predominantly) V-framed language (Wessel-Tolvig and Paggio, 2016), 2016), or English (S) and French (V) (Hickmann et al., 2011).

Regarding the semantic typology of motion event structure and gesture, Slavic languages were not addressed until recently. Kateřina Fibigerová and Michèle Guidetti (2018) compared production of French (V) and Czech (S) speakers from three different age groups in order to take the possible age effect[84] also into account. They examined elicited narratives of 144 subjects (48 subjects in each age group: 5- and 10-year-old children, adults) presented with videoclips of simple motion events which yielded more than 2 500 motion event constructions accompanied by gestures. Contrary to the line of evidence addressed above, Fibigerová and Guidetti did not observe the expected crosslinguistic effect. Both French and Czech subjects tended to gesturally mark only the PATH of the motion, not MANNER. Motion constructions with gestures reflecting both PATH and MANNER (although with a low absolute frequency) were significantly more frequent in adults, supporting the assumption that the complexity of gestural expression of motion events depends on developmental stage. As for the lack of the crosslinguistic effect, the authors offer two possible explanations. While not mutually exclusive, the two possibilities are (a) that, in terms of motoric effort, PATH gestures are easier to produce and therefore they prevail, sparing the speaker's motoric load, or (b) that the semantic component of PATH is cognitively more salient than MANNER, being more distant from the core of the conceptual domain of MOTION, and thus omitted from the gestural framing of the event. The fact that despite being proved to speak typical S- and V-languages, respectively, the way Czech and French speakers' gesture did not differ is, according to the authors, due to universal principles that underlie the gesture production, rather than language specific cognitive processes.

On the other hand, the authors admit that the occurrence of MANNER-gestures in the elicited narratives may be influenced by the salience of MANNER in the stimulus material itself. The incongruity with the previous research might actually arise exactly from this, suggesting that PATH is indeed generally a more prominent semantic component within the MOTION domain regardless of differences in linguistic encoding. The effect of language specific conceptualization reflected in gesture may thus exhibit only when MANNER is more relevant within the motion event semantic frame.

So far, analyses of elicited gesture production have brought relatively robust evidence of gestural encoding of event structure across languages, pointing to the dif-

[84] Suggested by previous studies (Allen et al., 2007).

ferent ways of "thinking for gesturing" – a phenomenon that complements Slobin's notion of *thinking for speaking*, comprising a single, interconnected cognitive mechanism, sharing the same conceptual basis.

4.2.2 Gesture and aspectuality

First analysis of the relation between gesture form and semantics of accompanying co-speech gestures was carried out by McNeill and Levy in their pioneering study (1982) that has served as a methodological baseline for many studies since then. McNeill and Levy analysed video-recordings of narratives elicited from six speakers. As a stimulus, a short cartoon was presented to participants who were subsequently recorded when retelling what they have seen to another person.

Gesture coding included gesture phases and multiple features of gesture form, as well as the type of a gesture (iconic, metaphoric, beat gestures). The authors' assumption was that co-speech gestures might manifest what McNeill had previously called the *concrete models* (see 3.2.1).

McNeill and Levy observed that the form of the iconic co-speech gestures coinciding with verbs correlated with verb semantics. These correlations included direct iconic mapping such as downward movement of a hand accompanying construction [V_{action} + *down*] but also the cases where iconicity was not so apparent. That is the case of Aktionsart features. In particular, McNeill and Levy found that the majority of gestures coinciding with the "end state" verbs (i.e. telic verbs or *achievements* in Vendler's (1967) classification) involved both hands, iconically distinguishing the two aspects of an action – movement of one hand depicts the process of reaching an end state, the other hand stands for the moment of reaching an end of an action.

As for the beat gestures, however, the authors did not observe any correlation between the formal features of the beat gestures and verb meanings. Majority of the beat gestures were used in the non-narrative context, whereas the narration itself was characterized by a presence of iconic and metaphoric gestures. From this discrepancy, the authors drew the conclusion that only the iconic gestures have the primary representational function. Beats, on the other hand, play a role in the structuring of discourse, being formally less diverse than iconics, they mark the boundaries of the parts of discourse. "Rather than control by the structure of the event being described, beats are controlled by other properties related to discourse structure" (McNeill and Levy, 1982, p. 287).

Although this study undoubtedly brought novel insights into the matter of co-speech gestures and has to be acknowledged as the initial impulse for a brand-new research area, there are several issues concerning the conclusions made by the authors that one should be aware of. First, the analysis was based on a sample of 145 gesture–verb combinations which is a rather small number with respect to the variety of verbal and gestural features that were included in the analysis.

Second, gesture production in elicited narrations may be rather specific compared to the gestures that occur in other communication situations that do not employ imagery to such an extent. It seems likely that when describing a plot of a cartoon, iconic gestures will be overrepresented in the speakers' production to the exclusion of beat gestures that may then appear as mere discourse markers that lack apparent iconic features.

Co-speech gestures in relation to eventuality were again addressed by Duncan (2002). In her research, Duncan compared co-speech gestures accompanying *perfective* and *imperfective* constructions in English and Mandarin. With respect to the aspectual distinction, comparison between these two languages is of course not straightforward. Whereas Mandarin discriminates grammatically between the two aspects, in English, (im)perfectivity is realized in terms of Aktionsart features of verbs and/or by the use of the progressive verb forms.

Duncan elicited narratives from 14 Mandarin and 11 English native speakers, adapting McNeill's and Levy's procedure. Besides the cartoon used by McNeill and Levy, Duncan also presented the participants with a series of short videos depicting animated figures engaged in various situations in motion and an hour-and-half long film.

Duncan extracted a sample of 100 combinations of co-speech gestures with (im)perfective verbs or constructions for each language (50 perfective and 50 imperfective) from the video-recordings of subjects describing what they had seen. The use of different stimuli sets enabled Duncan to elicit greater variety of gestures – particularly in the narrations based on the film plot that demanded subjects to employ more complex and abstract domains of conceptualization. This study exceeds the limitations of McNeill's and Levy's research also with respect to gesture coding – due to availability of more advanced tools for video analysis, Duncan was able to perform precise measurements of the duration of gestural units and thus to include a continuous variable applicable to all gestures.

Regardless of the gesture type – and even regardless of language, the results showed that the imperfective constructions are associated with (significantly) longer gestures in terms of duration when compared to those accompanying perfective constructions. "Imperfective" gestures were also observed to involve more complex or repeated movement patterns. This iconic mapping between gesture form and meaning of a construction, importantly, occurred across the scale of the affordance of the verbal meanings to iconicity.[85] Duncan thus concludes that the aspectual distinction may be one of the triggering factors in iconic gesture production. Moreover, the absence of significant differences between the two languages in the realization of (im)perfectivity-related gestures points to a fundamental role of aspectual opposition in cognitive processing that is not language-specific.

[85] See discussion in Chapter 6.

Another study of eventuality and co-speech gestures in English was carried out by Becker et al. (2011) again using data from elicited narratives. Elicitation procedure was in this case different from the two above addressed studies. Instead of video-stimuli, free narratives were elicited from pairs of participants who were asked to discuss a difficult or unusual situation they experienced. In total, the sample contained narratives of 8 speakers (69 minutes of recordings) – some of them, however, were not native English speakers. The authors did not report the level of English proficiency for the individual participants.

For the analysis, only the combinations of co-speech gestures with verbs in the past tense were selected (80 instances). In order to assess the possible correlation between gestural features and eventuality type, verbs were coded for telicity, durativity, and dynamicity, each gesture hence falling under one of the Aktionsart categories of Vendler's classification (*activities*, *states*, *accomplishments*, *achievements*).

The achievement verbs (i.e. dynamic, telic, and punctual, referring to an action or event that has reached its final state in an instantaneous manner) were found to be accompanied by gestures bearing a feature of punctual ending, whereas activity verbs (i.e. dynamic, atelic, and durative, referring typically to an unbounded action) were observed to be often associated with repeated or prolonged gestures.

Although such a limited sample did not allow for any kind of a statistical assessment of the possible correlations, the observed tendencies appear to be in accord with findings of Duncan (2002), as the perfective constructions often semantically intersect with achievements (and so do imperfective constructions and activities).

The authors subsequently carried out a comprehension experiment to investigate whether the gesture form contributes to a comprehension of eventuality constructions. They presented the participants (distinct a group from the elicitation task, again some of them non-native speakers of English) with extracts from the narration video corpus from the elicitation task. The extracts depicted production of a gesture – with either the punctual "achievement" features or durative "action/accomplishment" features. Half of the stimuli were combined with the original sound and the other half was combined with mismatching utterances (e.g. activity verb vs. achievement gesture). After viewing an extract, a verb (correct or false) was displayed on the screen and the participants were asked to decide whether the displayed verb is the same as the one they had heard.

In the mismatching condition, subjects' reaction times were significantly slower than when the video and audio channels matched. This suggests that the multimodal integration of verbal and nonverbal content also plays an important role in comprehension of eventuality constructions. "Aktionsart is not merely a grammatical distinction, but that these categories have cognitive reality in terms of imagistic construal of events and the mental simulation of them, and the grounding of language in action" (Becker et al., 2011, n.p.).

The results of this study support the idea of general tendency of co-speech ges-

tures accompanying the expressions of eventuality to iconically manifest the conceptual representation, as was suggested by the above addressed study by Duncan. Nevertheless, it has to be emphasized that also in the case of this study, the results are based on a rather limited sample and, also, that the possible effects of the various linguistic backgrounds of the participants are not reflected.

Another contribution to the exploration of the correspondences between co-speech gestures and the expression of eventuality in English was made by Parrill, Bergen and Lichtenstein (2013). In their study, 32 English native speakers were asked to read a short story on a computer screen. The story contained either verbs in progressive or present perfect form (there were two versions, participants were divided into two groups according to the version presented to them). Afterwards, participants were asked to retell the story to a partner.

The results revealed a significant tendency of the speakers to use longer (in terms of gesture duration) and more complex (repeated) gestures when producing verbs in the progressive form. However, this tendency was only present in such cases where they produced verbs that had been presented to them in progressive form in the stimulus story in the first place. According to the authors, this suggests that the iconic link which may have the function of embodiment of the corresponding concept in the speakers' mental representation, might be associated with situations where the speaker is forced to focus on the internal structure of the event:

> "The fact that a speaker uses one aspect or another does not necessarily tell us anything definitive about their conceptualization of the described event. Our results showed that production of the progressive goes along with increased focus on event internal structure, but only when we know that the speaker learned about the event in such a way that he or she was encouraged to encode the event-internal structure" (Parrill et al., 2013, p. 154).

These findings are interesting not only because the progressive aspect – and not the perfective – is associated with a specific gesture form, suggesting some kind of "markedness" of the progressive, but also concerning the role of the on-line language processing in the choice of the particular gesture form. The influence of the recent experience with comprehension of the progressive verbs on the production of multimodal utterance suggests that speakers may indeed draw more attention to the internal structure of the event – facilitated also by the possible "markedness" of the progressive in English – which may lead to the activation of the associated image schema. Also, this provides support for the *thinking for speaking* hypothesis.

As of today, the most extensive (in terms of number of languages captured as well as absolute number of speakers involved) study on gesture and eventuality is a cross-linguistic survey carried out by a team led by Alan Cienki (Cienki and Iriskhanova, 2018). Cienki's group investigated the relation between aspectuality and gesture in three languages: French, German and Russian.

Taking the cognitivist stance, they approach aspectuality in terms of the image schemata (2.2) of BOUNDARY, linguistically represented in aspectuality categories, and embodied in gestural patterns (with a boundary on the onset, offset or both, or with multiple boundaries (Müller, 1998). Schemata thus allowed the authors analyse the event construals encoded in grammatical aspectual forms and lexicalized Aktionsarten alike. Inspired by Sasse (2002) the authors distinguish *event unboundedness* (lacking boundaries, encoded via imperfective aspect or atelic Aktionsarten such as gnomic processes or states) in contrast to *event boundedness* (with a variable boundary complexity, realized linguistically in perfective aspect or telic Aktionsarten, e.g. punctual activities).

This distinction is understood as a general cognitive process – whenever an event is construed, the speaker carries out a binary operation of eventual perspectivization: choosing between bounded or unbounded construal. Given that, the authors can disregard the crosslinguistic differences in their language sample, as long as the languages contain morphological or semantic devices that are grounded in this general dichotomy:

> "understanding aspectuality as the cognitive basis for lexical and grammatical aspect provides us with the possibility to analyze aspect through the bi-dimensional model of perfectivity vs. imperfectivity, and, in addition, to take into consideration various manifestations of aspectuality in different languages – grammatical (tenses), lexico-grammatical (morphology), and relations to pragmatics" (p. 49).

Throughout their study, the authors sometimes refer to the linguistic representation of BOUNDEDNESS as (im-)perfectivity, as it is purportedly grounded in verb systems of the three languages (p. 61). This does seem to hold even at the most general level, though. If we postulate a theoretical comparative concept (Haspelmath, 2010) of aspect with two values – *perfective* and *imperfective* – and vaguely defined in terms of boundedness, we are faced with the fact that, in the three languages in question, the overt aspectual encoding in verbs is realized in a rather different manner and to a varying degree. Even in Russian, a language with grammaticalized aspect, the aspectual distinction does not apply across all tenses. In French, (im-)perfectivity is encoded in verb forms only in the past tense, whereas in German, verb morphology does not convey aspectual meaning, despite the terms *Perfekt* and *Imperfekt* (*Präteritum*), denoting the two past tense forms. In German, the two forms are highly lexicalised: some verbs are almost exclusively used in one form, otherwise the choice of the past tense form is a matter of register or narrative viewpoint.

Crucially, compared to French and Russian, the German aspectual system is highly variable across the many dialects. As far as Standard German (Hochdeutsch) is concerned, it can be compared to French when it comes to the modality-specific distribution of different past tense forms: in spoken German, Perfekt dominates for most verbs (except for the high-frequency modal verbs) that, if used in Imperfekt would

be characteristic of written or formal registers. In spoken French, the analytic *passé composé* is preferred form of the "perfective" past tense, whereas *passé simple* is characteristic of writing.

Compared to German and French that exhibit a higher degree of lexical-semantic (Aktionsart) restrictions, the Russian system allows for most of the verbs to occur in either form.

I disagree with the authors in the assumption that "French, German, and Russian share two broad aspect(ual) categories in their verb systems" (p. 61) as it is true only terminologically speaking. The difference between the two German past tenses, although it is described as aspectual in German grammars (p. 91), does not seem to fit within the same conceptual domain (boundary schema) as Russian and French (despite the mutual differences) do – or, in other words, regardless of whether the imperfectivity/unboundedness vs. perfectivity/boundedness distinction is actually relevant in the event construal in German speakers, it definitely does not "directly" correlate with the distribution of *Imperfekt* and *Perfekt* (see also Dahl, 1985, for a large-scale crosslinguistic survey of (im-)perfectivity marking).

An integral part of a boundary schema is a kinesic schema[86] – a visual manifestation of the embodied boundedness: "bounded movements are a gestural way to embody perfectivity while unbounded movements embody imperfectivity" (p. 57). But how can the gestural correlate to the linguistically encoded event boundedness be identified? The authors chose an approach to gesture boundedness that is form-based (i.e. disregarding the representational content of the gesture) and based not on phonological, but on *kinesiological* parameters (Boutet, 2010). The main parameter of interest considers the quality of movement, labelled as a *pulse of energy*. Bounded gestures are characterized by an identifiable pulse of energy (or "pulse of effort")[87] "marked by a tension in the gesture, through a kinematic form, through accelerations, jerks or stops, or very controlled tension of the muscles opposing the movement, which can be seen thanks to subtle changes in the movement" (p. 108). Unbounded gestures, on the other hand, do not exhibit a visible pulse of energy. Somewhat hard to define, as it is, the pulse-of-energy criterion is also approximated via the notion of *movement control* (or "gain control").

The crosslinguistic study consisted of a corpus analysis and a gesture perception experiment. The corpus part was based on a collection of six hours (two hours for each language) of recordings of semi-spontaneous production, elicited using the same method as in the study by Becker et al. (2011). In total, 80 participants were recorded (20 German, 22 French and 36 Russian speakers), producing 1160 gesture phrases in total, relatively evenly distributed across languages (38% German, 36% Russian, 27% French).

Only the French data supported the assumptions of multimodal expression of

[86] In this sense, BOUNDARY may also be viewed from the force dynamic perspective (2.2.)

[87] The notion of *effort* was inspired by the Labanian movement analysis (Laban and Lawrence, 1947).

BOUNDARY schemata: 71% of gestures accompanying the *passé composé* were bounded, whereas unbounded gestures accompanied 67% of the verbs in *imparfait*. In German and Russian, however, bounded gestures predominated with both *Imperfekt* and *Perfekt* and imperfective and perfective, respectively. Both in German and Russian, the proportion of bounded gestures accompanying the *Perfekt*/perfective was greater that in the case of *Imperfekt*/imperfective, but only in Russian the difference between proportions of bounded gestures with the two aspects was significant.

The methods of the experimental part of the study were adopted from the study by Becker et al. (2011) as well (i.e. a verb recognition task was used). A total of 161 subjects (54 French, 56 German and 52 Russian) were presented with video clips selected from the material used in the corpus study. The video clips captured speaker producing bounded or unbounded gestures with either original video or were edited to mismatch the verbal and gestural expression of the BOUNDARY schema.

An overall tendency across the three languages to faster responses in the bounded condition (i.e. perfective + bounded gesture) than in the mismatched condition perfective + unbounded gesture. In both matched and mismatched conditions with speech, the reaction times differences were not significant. The authors[88] suggest that the lack of an effect in the latter case is related to the the crosslinguistic differences in the IMPERFECTIVE domain, that appears to subsume a rather diverse categories, whereas in the case of the crosslinguistic concept of perfective in this study can be justified because "as morphological markers, the verb forms for these categories are similar semantic cues for guiding a "perfective" mental simulation of an event" (p. 176).

Based on the results of both the corpus-based and experimental, the authors stopped short of calling the construal of aspectuality multimodal *per se*. Rather, they considered the degree of gestural co-expression of aspectuality to follow language-specific patterns related to the encoding of event structure in grammar. Specifically, according to Cienki et al., the systematic use of specific gesture forms in the construal of aspectuality could be associated "[with] achieving the balance between more abstract grammatical meanings and more specific lexical meanings" (p. 180). This would explain the difference between French where aspectuality is encoded in the tense-aspect system and German where aspectuality is more of a lexical category. The authors assume that the same holds for Russian, in which aspect is "a lexico-morphological category that manifests itself both at the grammatical and lexical-semantic levels" (p. 181).

The most recent contribution to the study of multimodal expression of eventuality is provided by a paper by Jennifer Hinnell (2018). Subscribing to the multimodal CxG stance, Hinnell analysed 250 instances of five English verbs gathered from the Red Hen archive. She focused on auxiliary periphrastic constructions such as *stop making up stories* or *continue to grow* where the verbs *stop* and *continue* are grammaticalized into aspectual[89] markers.

[88] The experimental part of the study was conducted by Ray Becker and Monica Gonzalez Marquez.

[89] According to Levin (1993), the verbs in the constructions analysed by Hinnell belong to the class

Hinnell works with the distinction between *open* and *phase aspects* proposed by Frawley (1992) as typological macro-categories subsuming various aspects, both lexical and grammatical. Open aspects "express[es] extension of an event over a time frame" whereas phase aspects reflect "how [events] change status inside or outside the time frame" (p. 328).[90] In this semantic typology of aspect, languages differ in the tendencies in profiling specific aspectual types while underspecifying others. English belongs to "open languages" as it puts more semantic restrictions on the progressive aspect compared to other forms that are defined more vaguely (*ibid*., 330).

Importantly, Hinnell analysed only the usage of the target constructions occurring in "naturalistic and interactional contexts" (p. 10), while focusing on "parameters traditionally considered to be the loci of meaningful content in gesture [including] stroke segmentation into action phases and gesture onset timing in an examination of aspectual constructions" (p. 31). For the gesture analysis, she used Bressem's (2013) LASG system (see 4.3.2), proving the general suitability of the sign language phonology-derived categories for operationalizing the description of gestural behaviour.

From the 250 instances, 147 constructs was accompanied by gesture – most frequently so in the case of [*continue + to V*] (74%), followed by [*keep + V*] and [*quit + V*] (both 58%), [*stop + V*] (54%) and [*start* + V] (50%). Apart from relative frequency of gesture-accompanied instances of the target constructions and customary association statistics (χ^2 test), Hinnell does not discuss the degree of constructionalization, leaving this issue aside without an ambition to contribute to the debate of how to assess it quantitatively (see the discussion in Section 3.2.3).[91]

The strength of Hinnell's approach is in the qualitative description of what she aptly calls *aspectual contours* describing the inner force-dynamic profile of an event – an approach akin to one introduced by William Croft in his aspectual typology (2012, discussed below), although without a reference to his work.

What stands out from the findings is that, on the most general level, Hinnell observed that the "open aspect" constructions (i.e., [*continue + to V*] and [*keep + V*] were accompanied by complex gesture phrases, while simple gesture phrases were more strongly associated with "phase aspect" constructions [*start + V*], [*stop + V*] and [*quit + V*]). Focusing on the number of "action phases", i.e. movement variations within the stroke phase (beats of the hand or individual circles of a cyclic gesture) corresponding

of *Aspectual verbs* (subclasses of *Begin verbs* and *Complete verbs* §55.1, §55.2), cf. also *Aspectualizers* (Freed, 1979). Such terms, however, do not seem to be felicitous, as they cover only one section of aspectuality domain. More fitting term would be *Phase verb* used widely in Slavic aspectology.

[90] The system also comprises the third category of closed aspects that express "restriction of an event to a time frame" (p. 328) - typically represented by perfective aspect or punctual Aktionsarten.

[91] Regarding the assessment of constructionalization, Hinnell assumes that "[w]hile many more studies are required to investigate degrees of entrenchment, the very fact that conventionalization between gesture and linguistic form can be shown suggests that these co-speech gestures ought to be considered a part of inherently multimodal constructions" (p. 31).

to the segments distinguishable (i.e., profiled) within an event contour.

Distinct clusters of gesture forms were associated with the respective constructions, corresponding to Aktionsarten as well as other semantic features not only of the auxiliary verbs as such, but also beyond lexical level: it is in fact not particularly felicitous to talk about aspect here, as both open aspect and phase constructions consist of inherently progressive gerund forms (except for *continue* that mostly occurs with the main verb in infinitive form). For the sake of this work, let us focus only on the gesture phrase structure reflecting the general shape of an event contour.[92]

Open and phase aspects are defined by the presence or absence of event structure boundary profiling. Although one can be led to a conclusion that the single gesture stroke is what profiles the boundary in the phase aspect constructions, whereas the complex gesture phrase marks the complexity of a protracted action expressed by the open aspect constructions, the qualitative account provided by Hinnell does not quite support such an inference. The link between a complex contour and multiplicity of action phases seems to hold, but in case of phase aspect constructions, the tendency to a single stroke cannot reliably be interpreted as marking the onset/offset of an event, but rather a lack thereof. An exemption may be the verb *quit* having a prototypical gestural profile with "a relatively higher velocity (qualitatively observed) and an abrupt end to the stroke phase" (p. 20).

Gestural marking of eventuality was also addressed by Lis and Navaretta (2013) who measured the predictability of the form of gestures co-occurring according to eventuality category of the accompanied verbs in a multimodal corpus of Polish.[93] The classification of verbs was automatically extracted from the Polish version of *WordNet* lexical database (pl.WordNet 2.0).[94] *Eventualities* here refer to both Aktionsart categories and aspect as well as the category *eventuality type*, which collapses the eventuality-related semantic domains from WordNet database into two macro-categories: *translocation* and *body motion*.

Annotation of gesture form followed the PCNC schema (Lis, 2014) based on four phonological features (handshape, movement, direction and location) and additional features such as viewpoint (OVPT, CVPT and dual viewpoint), or mode of representation

[92] Left unmentioned here are also Hinnell's findings concerning the differences in gesture onset timing across the two aspectual constructions, which is a novel and potentially important observation: according to Hinnell's study, the shift between the gesture onset and onset of the corresponding lexical unit is significantly shorter in the phase aspect construction. This difference is interpreted as a potentially iconic reflection of a force-dynamic contrast. However, the reviewed study is not explicit about how the speech onset was captured. Also, the shift depends on other variables including the duration of the lexical unit (Jehlička, 2016). Further exploration in this regard would be by all means interesting, as the length of the word might be one of the factors contributing to the overall construed iconicity (cf. the iconicity principle of quantity, Givón, 1991): consider the length of *start*, *stop* and *quit* vs. *continue*.

[93] *The Polish Cartoon Narration Corpus* (PCNC, Karpiński et al., 2008) consisting of roughly an hour of recordings of 10 speakers re-telling the Canary Row film.

[94] http://plwordnet.pwr.wroc.pl/wordnet/

(depicting, acting indexing or embodying).

Lis and Navaretta analysed the association between the gestural features and eventualities on the sample of 269 multimodal pairs. They ran a series of machine learning experiments to assess how precisely can the algorithm predict the occurrence of the annotated gesture formal features from the eventuality categories of the verbs. The results showed two tendencies: first, that the distinction between translocation and body motion eventuality types was a relative strong predictor of the gestural viewpoints, and second, that Aktionsart, to a lesser degree, contributes to the prediction of the phonological feature of direction. Aspect, on the other hand, did not correlate with the gestural forms at all.

The modest sample size limits the generalizability of the results of this study. However, it still is noteworthy as it introduced a promising methodological framework for further investigations into the relation between gesture form and eventuality that could be employed on larger samples, using the eventuality classification of verbs available for many languages in the WordNet database.

Compared to the domain of MOTION the multimodal expression of eventuality has received a relatively lesser attention. Apart from the small-scale study by Lis and Navaretta (2012), three gestural formal features have been in focus in the multimodal studies of eventuality. One of them was the temporal duration of gesture stroke. As I already pointed out above, the duration of a gesture is affected by a wide range of factors including the duration of the accompanied lexical unit and as such requires a fine-grained phonetic analysis of speech that is beyond the scope of this study and thus the issue of the gesture duration will not be addressed here. The other two formal parameters were the complexity of a gesture phrase and gestural marking of an event BOUNDARY.

The accumulated evidence converges towards an observation that the *complex* types of gestures (i.e gestures with salient movement modulations within the gesture stroke or by repetition of the stroke) in general appear to be linked to *imperfective* or *open* aspects, in particular, they seem to be associated with the English *progressive* verb forms. Gestures that exhibit some kind of BOUNDARY marking have been associated with *perfectivity* in general, or to Aktionsarten that either highlight the event's ending or its punctual nature.

The picture that emerges is that across languages, event BOUNDEDNESS may be embodied in the forms of co-speech gestures that may be consider conventional to a certain degree, whereas the unbounded events *per se* do not seem to be systematically associated with specific gestural forms. However, some types of unbounded events, particularly those with lexicalized verbal representations (such as the English progressive), may be associated with recurrent formal features (such as various types of complex gestures).

The crosslinguistic study by Cienki and Iriskhanova (2018) has been the most ambitious venture into the domain of multimodal expression so far. Yet, it has also

brought forth two potential methodological caveats that need to be dealt with in this study.

The first issue concerns the coding of the linguistic expression of event boundedness. Cienki's and Iriskhanova's coding was based simply on the aspectual forms available in the given languages. Such approach is problematic for two main reasons. First, categories such as *perfective* are not universal, e.g. one language's perfective never entirely overlaps with another language's perfective semantically and in terms of functional distribution. These are what Haspelmath (2010) called *descriptive categories*, which are language-particular. For the purposes of a comparative analysis, one should postulate language-independent semantic parameters (corresponding to the underlying conceptual representations). Thus, an *onomasiological* approach (see Geeraerts, 2010, for a cognitive perspective on onomasiology) to linguistic encoding of eventuality will be proposed here (4.3.1), following the tenets of CL programme.

Second, grammaticalized aspect is just one of the means of the linguistic encoding of event BOUNDEDNESS. An assumption can be made about the constraints on event conceptualisation related to the presence or absence of grammaticalized aspect in the given language, but in order to investigate the multimodal construal of event boundedness, one needs to take into account the *lexical semantic* properties of the verbs under scrutiny, as well the semantic properties of the constituents within the verb phrase. Finally, an account of multimodal event construals would not be complete without considering pragmatic, inter-subjective and socio-cognitive aspects of the communicative situation in which they take place (Schmid, 2016).

The second problem concerns the coding of gestural representation of BOUNDARY schemata. While approaching gestural boundedness in terms of kinetic properties is in principle legitimate, the question is whether it is the best way regarding the operationalization for human coders: it is not at all clear if it is possible to establish a reliable visual discrimination of an "energy pulse". An alternative way will be proposed here (Section 4.3.2), based on formal features inspired by the well-defined and established parameters of sign language phonology (Stokoe, 1960).

4.2.3 Expression of eventuality in sign languages

With regard to gestural marking of eventuality in spoken languages, none or only little attention has been paid to how the same semantic distinctions are expressed in sign languages. However, taking a closer look at languages primarily realized in the visual-motoric modality may provide important insights into general and modality-specific aspects of linguistic representation of event structure.

As do spoken languages, sign languages rely on a variety of means of the expression eventuality. These include lexical-semantic properties of signs (Aktionsart) on the one hand, and aspectual marking via grammaticalized lexical units (such as FINISH or HAPPEN in ASL) or phonological modification of the sign on the other. Concern-

ing the latter, Ronnie Wilbur focused on how event structure may be reflected in the phonological structure of predicates (non-classifier as well as certain types of classifier predicates) in ASL and beyond (2003; 2008). Wilbur's work inspired a further research carried out under the flag of the *Event Visibility Hypothesis*.

Initially focusing on the expression of telicity in ASL, Wilbur described that certain phonological configurations occur as stable representations of event structure. In ASL, predicated construed as telic (corresponding to Vendler's accomplishments and achievements), are characterised by a set of movement types, all sharing the phonological component of an abrupt halt of the movement (Figure 15),[95] while the atelic predicates do not exhibit such property. The same pattern was observed also in Austrian Sign Language (ÖGS, Schalber, 2006) and Croatian Sign Language (HZJ, Malaia and Wilbur, 2012).

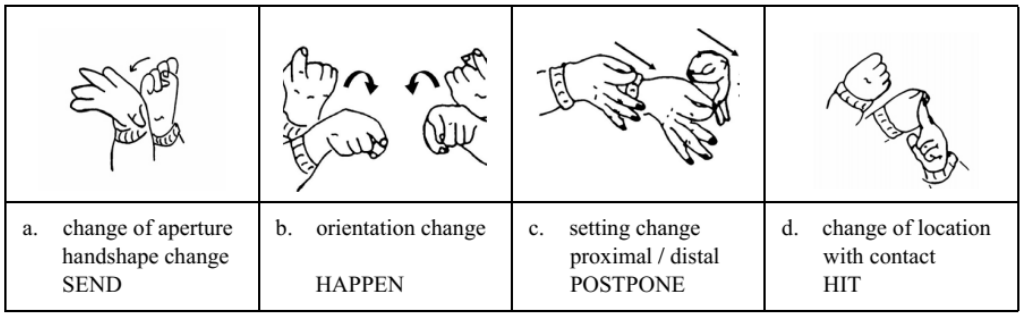


Figure 15: *Movement types in telic signs*

Testing the hypothesis that "sign languages denote telicity by perceptual 'end-marking,' as potentially measured by the slope of deceleration from peak velocity to the end of the sign" Malaia and Wilbur (2012, p. 128) measured the slope of deceleration in telic and atelic predicates produced by native HZJ and ASL signers, recorded by a motion capture device. They found that in both languages, the telicity is indeed marked by an increased movement velocity followed by a rapid deceleration. In HZJ, moreover, the variation in these phonological parameters constitute minimal pairs: a single handshape (lexical root) may be construed as telic or atelic by phonological modulation. This typological difference to ASL is likely to be a matter of contact between HZJ and the dominant spoken Croatian that marks verb aspect morphologically.

In a series of experiments, Strickland et al. (2015) presented hearing non-signers (native speakers of English) with videos of telic and atelic signs produced by native signers of Turkish Sign Language (TID), Dutch Sign Language (NGT) and Italian Sign Language (LIS). The subjects' task was to choose the correct meaning of the sign from two options. In two of the experimental conditions (with LIS stimuli), one of the options was the target meaning, whereas the other, with different telicity, was from either the same or different semantic domain. In another condition (ran separately with TID,

[95] Source of the image: Wilbur, 2008, p. 232.

LIS and NGT stimuli), neither option corresponded to the stimulus meaning, but one the meanings matched the stimulus' telicity. The authors also presented one subject group with non-signs artificially created to mimic telic and atelic signs, with the former having "an abrupt stop in movement, contact, and/or a sudden change of hand shape" and the latter "rapid, repeated motion or 'trilled movement' [...] that lacks a salient gestural boundary at the end" (Strickland et al., 2015, p. 5970). In all conditions, the subjects were able to guess the correct meaning or to choose the appropriate telicity value with a significant accuracy. Furthermore, in the subsequent experiment, the stimuli from the three sign languages as well as the non-signs were rated by non-signers in terms of the presence of a boundary or repeated movement. The rating study revealed that the non-signers can perceive the formal distinction between telic and atelic signs, suggesting that subjects in the behavioural experiments were indeed guided by the perceived "visible telicity".

Some argue (Strickland et al., 2015; Kuhn, 2017) that the gestural/signed representation of boundaries is a manifestation of universal cognitive mechanism organizing the human perception of events. Further exploration is necessary, however, to tease apart (or, rather, to understand the entanglement of) the underlying cognitive constrains on perception from the constrains that arise from habitual linguistic patterns (i.e. the further research should involve subject groups with more diverse linguistic and cultural backgrounds).

4.3 The present study

This study deals with the relation between gestural form and the linguistic encoding of eventuality in English and Czech. Both languages have been already represented in the study of the relation between gesture and the expression of eventuality. However, the two languages have not yet been compared directly in this regard. The use of gesture by the speakers of Czech was only investigated in the domain of MOTION (Fibigerová and Guidetti, 2018).

The present study was carried out in two stages: first, a quantitative analysis of the association between linguistic and gestural encoding of eventuality was conducted using two samples of naturalistic production of English and Czech speakers (reported in Chapter 5). Subsequently, a behavioural experiment was run with the Czech subjects only, testing whether some of the patterns observed in the production part of the study are also involved in the comprehension of multimodal utterances.

Adopting both the production and comprehension perspectives on the multimodal construals of eventuality, this study follows in the research line represented by the studies by Becker et al. (2011) and Cienki and Iriskhanova (2018).

This study presents a novel approach that should overcome the issues of the previous studies mentioned above. An onomasiological approach is adopted to the

linguistic encoding of eventuality, drawing on Croft's (2012) model of event semantics (4.3.1). Analysis of gesture is based on the phonological features (4.3.2) and uses the parameters from the annotation system developed by Jana Bressem (LASG, *Linguistic Annotation System for Gestures*, 2013).

4.3.1 Cognitive model of event semantics

Considerable attention has been paid to BOUNDEDNESS as a semantic property not only with regard to verbs. According to Langacker (1987a; 1987b), *bounding*, i.e. profiling of a boundary within a semantic domain (applicable to the domains that are dimensional), is central to the distinction between perfective and imperfective verbs: perfective processes are "bounded in time within the scope of predication" (1987b, p. 80), whereas the imperfectives are boundless in this respect. There is a conceptual affinity between aspectuality and the un-/countability of nouns: „the region profiled by a count noun is specifically bounded within its primary domain" (*ibid.*), whereas mass nouns are characterised as not having a specified boundary. The verbs are thus considered a specific case of bounding within the domain of TIME.

Janda (2003) suggested analysing the (Slavic) aspect in terms of a properties of matter rather than time, based on a type of the TIME IS SPACE metaphor:

> "What is there about boundedness, totality, definiteness, resultativeness, exterior vs. interior, figure vs. ground, and punctuality vs. durativity that makes all these concepts hold together in the meaning of aspect? Answer: They are the properties of matter that serve as the source domain for the metaphorical grammatical category of aspect" (Janda, 2003, p. 252).

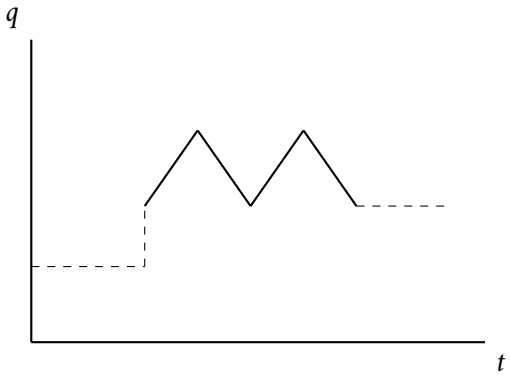
The Russian perfective is viewed as a "discrete solid" and the imperfective as a "fluid substance". Such a view entails a multiplicity of qualities of matter that may be profiled in the use of a specific aspectual form – not only the boundary of temporal unit, but also the inner qualities of the objectified item, such as its shape, structure or consistency. Looking at aspect from the perspective of this conceptual metaphor, one may naturally assume that the boundaries of metaphorical discrete solid entity or the inner qualities of the boundless fluid substance will be embodied and thus made visible in gesture.

In his model of aspectual types, William Croft (2012; 2016) adopts a two-dimensional view of boundedness. Croft's model elaborates on the canonical Vendler's distinction of the four situational types: *achievements* (ACH), *accomplishments* (ACC), *activities* (ACT) and states (STATE). However, Croft expects the whole system of aspectual types to be inherently more flexible, allowing for dynamic conventionalization and creative meaning extensions according to the communication needs of the community, within the structural limits of a specific language. Such flexibility stems from the distinction between verbs as lexical items on the one hand, and on the other, particular aspectual types as image schemata conceptualizing the prototypical *aspectual contours*

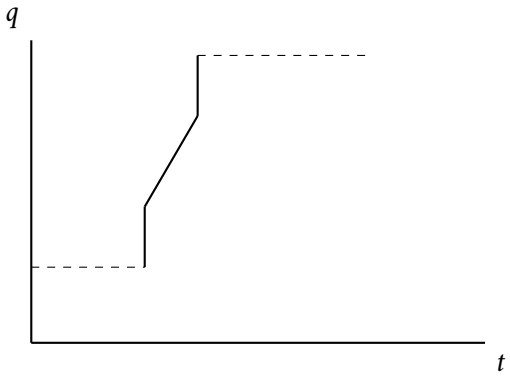
(see below) of events that we experience. Thus, a core verbal meaning may be realized through various aspectual types, given a particular situation, with the range of variance depending on the aspectual potential of the verb.

Aspectual types are determined by their aspectual contours. To specify their properties, Croft (extending Parsons' (1980) phasal model based on temporal boundedness) develops a phasal model of aspectual contours that unwind along two axes: temporal (*t*) and qualitative (*q*), referring to the temporal development and to the inherent structure of an event, respectively. A *t*-bounded aspectual contour is limited on the temporal axis; a *q*-bounded one has a closure on the qualitative axis. Each aspectual contour profiles a particular phase (or phases) as FIGURE and leaves some phases unprofiled (either as the presuppositions or the implications of an event). The phasal model allows for assigning each aspectual contour a bi-dimensional representation. Let us illustrate the bi-dimensional representation of aspectual contours in examples 10 and 11.

(10) *Yesterday, she was editing the video from 3 to 5 pm*



(11) *He baked a cake for his family yesterday*



Example 10 represents a directed ACT, example 11 represents a directed ACC. In the former case, it is the process of editing (a video) that is profiled (solid line), taking a dedicated amount of time and consisting of the repeated activity of editing subparts of the video. The initial phase of the process, although presupposed, is not construed by this aspectual type, and thus remains unprofiled in the contour. The same applies for

the (implied) natural endpoint of the event of editing a video, namely the completion of editing all the subparts of the videotape. This contour is *t*-bounded, but not *q*-bounded. The latter contour, on the contrary, being both *t*- and *q*-bounded, profiles the initial and the final phase of the event as well, leaving only pre-existing and resulting states (a non-existent vs. baked cake) unprofiled.

Apart from distinguishing between the *t*- and *q*-boundedness of an event, Croft enriches the original Vendlerian system with other relevant features: first, *directedness* refers to the event pointing towards a natural endpoint, regardless of whether or not the endpoint is profiled in the contour. Directed events with an unprofiled endpoint are not *q*-bounded and belong to directed ACT. In contrast, directed events with a profiled endpoint are *q*-bounded and thus represent either ACC or ACC. Second, Croft integrates into his model the notion of *incrementality* (Dowty, 1991; Krifka, 1992). A subtype of directed types of events, incremental events develop towards a natural endpoint that can be mapped on one of the arguments (typically objects or subjects), as in *Johnny has built a tower of cubes*, where the progress of building can be tracked through the subparts (*cubes*) of the outcoming tower, incrementally added on the top of one another. Not all directed events are incremental: some events develop heterogeneously, lacking such a gradual development (*repair a computer*). Finally, for undirected events, it is often the case that they refer to some kind of *cyclicity* – in this respect, the distinction between cyclic ACH and cyclic ACT is a prominent one. Cyclic ACH are semelfactives (*wink*, *cough*, *hop once*) that profile a single iteration of the denoted event (more specifically, the onset and punctual hold phase of the event). The cyclic ACT predicates profile the repetitive development of these events (*wink*, *cough*, *hop several times*), either as an actual event or as a habit.

In this study, a simplified version of Croft's typology of aspectual types that included only two levels of categorization was used, leaving out some features (e.g. permanence/transition of STATE, (ir)reversibility of ACH, heterogeneous vs. cyclic development of events) that did not allow for a straightforward operationalization for purposes of the manual annotation. These features were thus not considered in the predictions for this study. Croft's model has already been adapted for the description of the Slavic aspectual classes to a certain extent, by Croft himself and others (Janda, 2015; Kokorniak, 2017. Lehečková (under review), adapted Croft's model for Czech). Table 2 presents the final version of the taxonomy used in this study.

4.3.2 Multimodal construals of event structure

The embodiment of boundedness within the event structure in gestures is assumed to be signalled by the presence of a *marked ending* or *halt* in the gestural movement. This approach is conceptually akin to how boundedness is treated in sign language phonology (see the above discussed Event Visibility Hypothesis. Thus, instead of the concept of the *burst of energy*, one can get along using established coding guidelines

Vendler class	aspectual (sub)type		characteristics
state	state		does not involve change on *q*-axis; holds for a certain time span (potentially also temporally unbounded on both sides) (*contain*, *be a teacher*)
activity	undirected		*q*-unbounded (no natural endpoint implied); the internal structure of the process can be construed either as cyclic (similar internal phases) or heterogeneous (dissimilar internal phases); (*bark*, *run*, *breathe*, *sing*); including inactive actions (*sit*, *lie*, *think*, *consider*)
	directed	incremental	*q*-unbounded, but a natural endpoint implied, not profiled; gradual development of an event (*running a mile*)
		non-incremental	*q*-unbounded, but a natural endpoint implied, not profiled, heterogeneous development of an event (*be looking for a solution*)
accomplishment	directed	incremental	*q*-bounded (a natural endpoint profiled), durative, gradual development of an event (*run a mile*)
		non-incremental	*q*-bounded (a natural endpoint profiled), durative, heterogeneous development of an event (*sort it out*)
achievement	cyclic		punctual/semelfactive, *q*-bounded change from one state to another point in time (extreme contraction on time axis) (*sneeze once*, *give a wink*)
	directed		*t*-/*q*-bounded, change within a limited *t*-phase (*push down*)

Table 2: *Simplified model of aspectual types (adapted from Croft, 2012)*

for gestures based on phonological rather than kinesiological features, particularly the guidelines designed by Bressem (2013), who lists *accentuated ending* as one of the values of the *quality of movement* parameter:

> "The aspect 'quality of movement' specifically addresses the markedness of movements. A movement is marked, if it stands out in relation to other movements because of a particular saliency regarding one of these qualitative features. For instance, in an "accentuated" movement, the endpoint of the motion is stressed, because the movement is carried out with more force. This rise in force leads to an increase in the intensity at the end of the movement execution" (Bressem, 2013, p. 1090).

BOUNDEDNESS is not the only semantic property of event structure that will be in focus here. Besides the *ending* of the movement, which is hypothesized to profile the OUTER or RIGHT BOUNDARY of an event, also the inner structure of the event is profiled gesturally. Let us then propose a compositional approach to gesture boundedness, combining outer and inner boundedness:

(i) *ended* (e) gestures are characterized by a visually discriminable, abrupt stop, whereas *continuous* (c) gestures progress gradually, without a marked halt or rapid deceleration of the movement (see the last frames in Figures 16 and 17);

(ii) *complex* gestures are characterized by internal phases marked by repeated movements (multiple strokes) that either "divide up" the event into bounded segments by a sequence of ended movements or introduce unbounded complexity to the event profile by continuous movements, as in a repeated cyclic gesture (Figure 17).

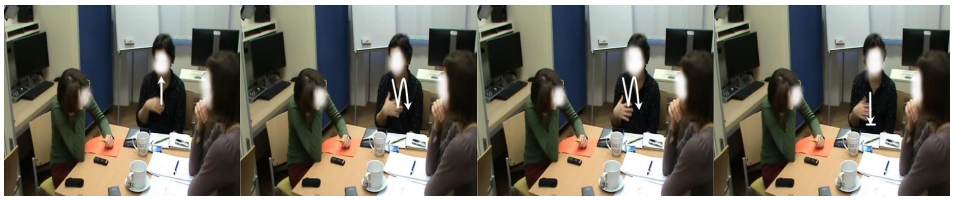


Figure 16: *Ended gesture with multiple ended phases (type = ee)*

Figure 17: *Continuous gesture with multiple continuous phases (type = cc)*

The modification of Croft's model together with the compositional approach to gesture boundedness allows for addressing not only the association between the event and gestural boundedness as such but also – and above all – to investigate potential gesture-speech integration at the level of event contours. The existing evidence points to the tendency of gestures to reflect the event structure of the concomitant linguistic expression discernible across multiple axes like perfectivity vs. imperfectivity or durativity vs. punctuality. One of the goals of the present study is to test the applicability of the aspectual contour framework to capturing the association between the linguistic and gestural encoding of eventuality.

Recall example 5 from Chapter 2. The gesture in focus is affiliated with the specific lexical unit – a part of the predicate, the participle form *rozfázovanou* ('divided up into phases'), which is an adjective derived from the verb *rozfázovat* and as such it inherits the verb's aspect (PFV). As the gestural form exhibits an apparent iconic mapping to the semantics of the participle, we may assume that the multimodal construct[96] is constituted by a combination of the "distributive" gesture and the participle and thus we can focus only on the aspectual type encoded in the participle form, which is a directed incremental ACC. The gesture is considered complex – it contains a multiplicity of movement subphases that are ended. Without an accentuated halt at the end of the complex movement, the gesture is considered continuous with respect to its outer boundedness. In this case, we can see a multimodal construal of the event's incrementality: the even distribution of the event's subparts (*phases*) is marked by the distributive prefix *roz-* and, at the same time it is highlighted visually by the iteration of the same movement pattern across the horizontal axis, splitting up the conceptual object (*tu praxi*, 'the training') into phases.

The aspectual contour of an event results from a construal operation, with gestures providing contextual cues for profiling the target part of the semantic frame. Depending on the situation, gestures have a variable weight in the profiling operation: sometimes they seem only to highlight the semantic information that is being profiled, sometimes it is (almost) exclusively the gesture that drives the profiling operation. The latter situation features gestures disambiguating potentially ambiguous sentences (Holle and Gunter, 2007) or providing critical information when other channels are not available (Drijvers et al., 2018).

These two types of the behaviour of gestures in construal will be referred to as *profiling-for-highlighting* and *profiling-for-discrimination*. Acknowledging this variability has crucial implications for the methodology of corpus data annotation (discussed in the following chapter).

[96] A *construct* is an instantiation of a construction (Traugott and Trousdale, 2013, p. 16).

4.3.3 Research questions and assumptions

The primary focus is on the possible mapping between *aspectual* and *gestural contours* of events. It is expected that in both languages, there is a tendency of the gestural forms to co-occur with certain event types systematically (in line with the Event Visibility Hypothesis). Besides, it is assumed that given the typological differences between English and Czech regarding eventuality expression, language-specific multimodal patterns will be observed. An apparent difference between the coding of eventuality in the two languages concerns the very notion of aspect: in Czech, as in many other Slavic languages, aspect is overtly marked by the verb's morphology, distinguishing between the *perfective* (PFV) and the *imperfective* (IPFV) via a productive system of derivational affixes (Comrie, 1976; Dickey, 2000). Most verbs thus come in the formally differentiated aspectual pairs (see examples 12 and 13).

(12) *∅-Psa-l-a* *jsem* *dopis*
IPFV-write-PST-F.SG COP letter

'I was writing a letter / I wrote the letter'. [in specific contexts]

(13) *Na-psa-l-a* *jsem* *dopis*
PFV-write-PST-F.SG COP letter

'I wrote/have written a letter'.

Note that although the term aspect is used here to denote the Czech PFV-IPFV distinction as well as the aspectual part of the English tense-aspect system (*simple*, *progressive*, *perfect*), we treat Czech and English aspects as language-specific categories. Aspectual types and subtypes are understood in terms of sets of semantic constraints on aspectual construals that may be encoded via morphological, morphosyntactic or lexical means in both languages alike. The following assumptions (not hypotheses in the technical sense, as this study is *exploratory*) can be postulated for the corpus part of this study on the grounds of the aggregated evidence from the previous studies:

(i) ACH and ACC (i.e the telic aspectual types) will tend to attract ended gestures in both languages;

(ii) the directed subtypes will also tend to attract ended gestures in both languages;

(iii) the incremental subtypes will, in both languages, tend to attract complex gestural forms;

(iv) the English PRG will be associated with continuous and/or complex gestures;

(v) in Czech, ended and continuous gestures will tend to cluster together with PFV and IPFV, respectively.

5. Corpus study

"Non solum corpus, sed aetiam spiritus!" (Už jsme doma: Ámen)

The first stage of the present study involves a corpus-based analysis of multimodal expression of eventuality in English and Czech. This chapter is organized in a way that is standard for reporting corpus research. The sources of the material are presented in Section 5.1, the description of the annotation schema, the operationalization of the coding and the assessment of annotation reliability follows in Section 5.2. Section 5.3 introduces the method applied in the quantitative data analysis and provides the descriptive statistics for the dataset. In Section 5.4, results of the quantitative analysis are presented (5.4.1 and 5.4.2), followed by a qualitative survey focused on interactional aspects (5.4.3). The chapter is concluded by an interim discussion (5.5).

5.1 Material

The previous findings of the multimodal corpus-based studies of eventuality (reviewed in 4.2.2) were supported by two types of data. Duncan (2002), Becker et al. (2011), Parrill et al. (2013), Lis and Navaretta (2013) and Cienki and Iriskhanova (2018) analysed gesture production occurring in elicited narratives and (in some cases) follow-up dyadic conversations between a test subject and a confederate. Hinnell (2018) used data from the Red Hen archive, specifically, she analysed televised talk shows, mostly interactions between the show's host and the guest.

Bearing in mind that "the core niche for language use in all cultures is a speech and gesture exchange system in which participants take short, rapidly alternating turns" (Levinson and Holler, 2014, p. 2), both narratives as well as television broadcasts are far from an ideal kind of data.

The repertoire of gestures that accompany the narratives elicited via visual stimuli (e.g. cartoons) may be considerably limited (there really are not many ways to describe a cartoon cat climbing up a drainpipe). This limitation can be overcome by using techniques such as asking the subjects about personal stories or issues like social dilemmas etc. (see Becker et al., 2011), especially when the confederate intervenes with follow-up questions, which may effectively induce a semi-spontaneous setting. Still, the subject's production is characterized by long segments and a relative lack of interactional phenomena (question sequences, repairs or gesturing for collaborative meaning-making).

TV recordings often involve technical issues such as voice-overs or cameras zooming in on only speakers' faces, as well as issues regarding the ecological validity of interactions, e.g. the presence of specific types of speakers whose gesture production

for some reason cannot be considered naturalistic (e.g. talk show hosts or politicians gesturing in "exaggerated" or "persuasive" manner), etc.

Although certainly more interactional than narratives and semi-spontaneous dyads recorded in laboratory settings, the TV recordings face us with issues that are beyond acceptable. Thus, one of the key challenges of this study a search for suitable multimodal corpus. For English, a relatively convenient source in this regard is the AMI corpus that includes recordings of spontaneous interactions of several speakers during business meetings, captured in highly ecologically valid conditions. For Czech, no such resource was available, therefore, a novel, special-purpose corpus had to be compiled. Leaving aside the additional time consumption, building the corpus from scratch was felicitous as it allowed for customization of the Czech sample so as it could be as comparable to its English counterpart as possible.

5.1.1 English subcorpus

The English production was sampled from the AMI corpus (Carletta, 2006), developed at the University of Edinburgh in 2004–2006. It consists of more than 100 hours of recordings of business meetings, with native as well as non-native participants (all meetings held in English). A larger part of the corpus consists of non-spontaneous sessions simulated under controlled conditions. A third of the material is represented by recordings of naturally occurring, spontaneous interactions. The corpus is equipped with multilevel annotation browsable in the NXT (NITE XML Toolkit) – an environment for linear transcription and annotation synchronized with the video and audio channels. Due to the projects' focus on the psychological aspects of multiparty interactions, the annotation captures discursive and pragmatic phenomena. Annotation of manual and head gestures is available only for the non-spontaneous part of the corpus. Speech is transcribed at the word level orthographically with a limited set of symbols for paralinguistic features (vocal noises).

From today's perspective, the AMI corpus, a project that is no longer in development, suffers from obsolescence on several levels: the software environment is dated, as are the technical specifications of the video recordings (low resolution).

On the other hand, it still has many major advantages. For a multimodal corpus, it is relatively robust, it is freely available, and the spontaneous part of recordings can be considered highly ecologically valid. For the purpose of the present study, only the non-scripted, spontaneous meetings were used.

Four sessions from the AMI corpus were selected as the English sample, featuring meetings of 3–4 people, mostly native speakers of English. The non-native speakers were excluded from the analysis (as well as one native speaker who produced al-

[96] AMI = *Augmented Multiparty Interaction*; http://groups.inf.ed.ac.uk/ami/corpus/.

[96] The transcription guidelines are available at: http://groups.inf.ed.ac.uk/ami/corpus/Guidelines/speech-transcription-manual.v1.2.pdf.

most no speech or gesture), leading to the final number of 9 speakers whose speech and gesture production was analysed (3 female and 6 male speakers with mean age of 29, ranging from 22 to 38 years). Each session lasted on average 46 minutes (with standard deviation of 9 minutes) and the total duration of the subcorpus is 3 hours and 4 minutes.

All speakers were either faculty members or students. Some of them were directly involved in the AMI corpus project, which is apparent from the meetings' agenda: topics discussed in three of four sessions are related to the AMI corpus; one session has no obvious connection to the corpus project as it features graduate students organizing a linguistics conference at the University of Edinburgh. The subject group consists of speakers of various English dialects: five speakers were from England, two were speakers of American English, one speaker was Scottish and one Canadian.

In three out of four sessions of the English subcorpus, there is a non-native speaker present (a different person in each session). Although this might be an important factor affecting the entire interaction and leading to the modulation of native speakers' speech as well as gesture directed towards the non-native speakers and possibly even to the other native speakers, it was not taken into account in the analysis. In all cases, the non-native speaker either interacted with others with ease, displaying high proficiency in English, or did not partake at all or very marginally, being practically left out of the interaction. Therefore, it is highly probable that the impact of their presence of the non-native speakers – even if it cannot be ruled out completely – was negligible.

5.1.2 Czech subcorpus

The Czech counterpart to the AMI corpus was built from scratch for the purpose of the present study. All recordings were made in Prague at the Faculty of Arts of Charles University with faculty members and students in 2016–2017. Some of the participants were (to various extent) aware of the multimodal corpus project, some were not. Such a heterogeneity within the participant group was intended as it made the Czech subcorpus perfectly comparable to the English subcorpus.

Four sessions were recorded with 2–5 speakers and the total number of 9 speakers (3 female, 6 male) one of which (female) appeared in all three sessions. All speakers were native speakers of Czech with an average age of 33 years (ranging between 25 and 44). Prior to recording, each speaker gave their consent with participating in the study (see informed consent form in Appendix B) and filled in a questionnaire (Appendix C) designed to collect participants' metadata (age, gender, region, language competence and handedness).

As in its English counterpart, the business meetings that were to be recorded

[96] Metadata available for speakers in the non-scripted part of the AMI corpus are limited compared to the scripted part.

were not simulated and they would have taken place regardless of the recording. The recording sessions were arranged with the meetings' convenors who were asked to relocate their meetings to a room equipped with video-recording technology in order to be captured on camera for the purpose of "the analysis of naturally produced speech in interaction". The focus on gesture was not revealed to the participants until the sessions were over.

Two sessions were recorded in a recording studio provided by the Institute of Deaf Studies of Charles University, one session took place in interpreting laboratory of the same institute and the fourth recording in an office where the meeting was originally scheduled. All four meetings were captured on two HD cameras (Sony Handycam DCR-HC51E and Panasonic HC-X920 recording in MPEG format with 1440×1080 px resolution and a frame rate of 25 fps) facing the table with mutual angular difference of ca. 50–70° (see Figure 18 for a schematic view of the recording setting).

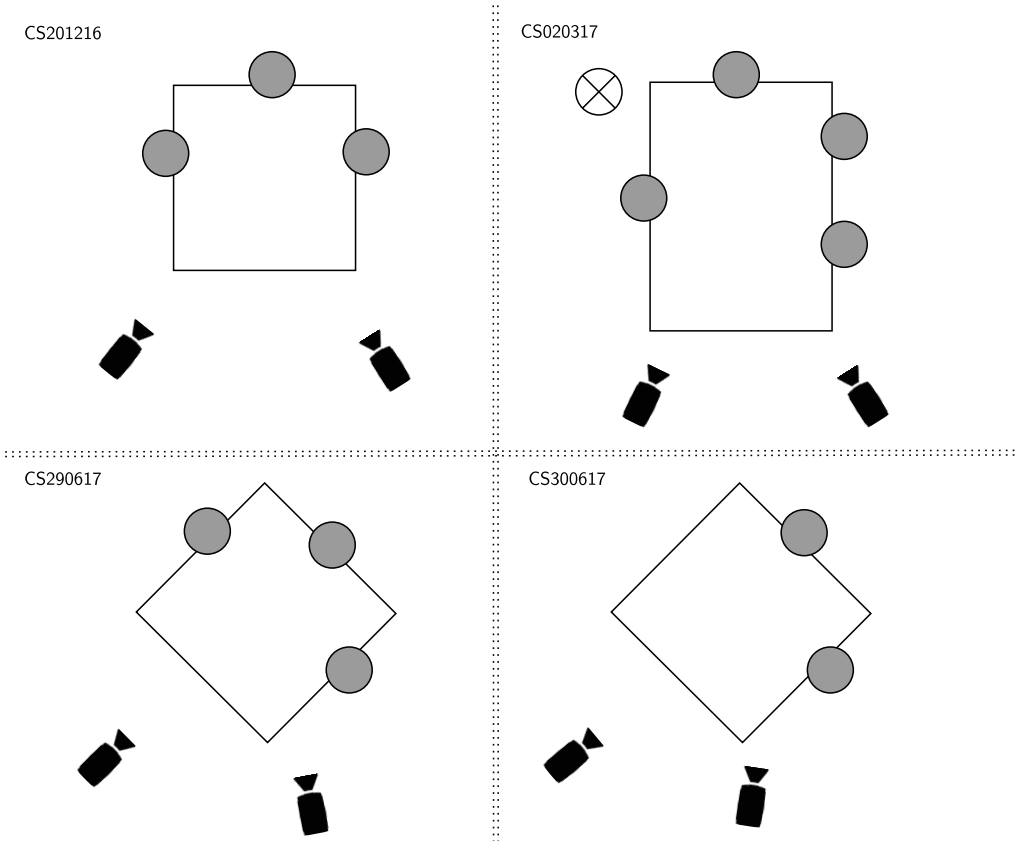


Figure 18: *Recording session schemata*.
Rectangles represent the table, grey circles represent the individual speakers, ⊗ in CS020317 designates the speaker that was not included in the analysis.

Although the Czech sub-corpus was designed to be as comparable to its English counterpart as possible, the two sub-corpora diverged in some respects – see Table 3 for the overview of divergences. Tables 4 and 5 provide an overview of the sessions included in the two subcorpora.

	English subcorpus	Czech subcorpus
Cameras	corner view, bird's eye view, zoom-in at participant faces	2 overview cameras
Sound	clean separate channel for each speaker (head-set microphones)	voice recorders set on the table
L2 speakers	present in some sessions	none
Recurrent participants	no	yes
No. of participants per meeting	3–4	2–5
Meeting duration	45 minutes – 1 hour	45 minutes – 1 hour
Topics	corpus building, academic agenda	corpus building, academic agenda
Setting	speakers seated aroud a table	speakers seated aroud a table
Genre	academic meeting (informal)	academic meeting (informal)

Table 3: *Comparison between the two subcorpora*

5.2 Annotation

The annotation was performed in ELAN, a software tool originally developed for the language documentation purposes but widely used for analyses of spoken and signed language, gesture and other bodily expression as well. ELAN is particularly well suitable for multimodal corpus-based studies, as it allows for time-aligned annotation of video and audio (with millisecond precision) on an unlimited number of levels and also offers a number of analytic tools including a search engine using regular expressions.

5.2.1 Transcription

As has been already mentioned above, the AMI corpus comes with a multi-layered annotation provided in XML format. Although using the original annotation software NXT did not turn out to be practical, the source XML files could be used for further processing. This was the case of the speech transcription files: it was possible to export the XML files to ELAN via PRAAT (Boersma, 2001) as separate tiers with individual words as annotation.

[96] ELAN = EUDICO (= *European Distributed Corpora Project*) *Linguistic Annotator*.

[96] ELAN was developed at the Max Planck Institute for Psycholinguistics in Nijmegen by Han Sloetjes and Peter Wittenburg (2006).

[96] Features of ELAN are subject to a continuous development, annotation for the present study has been carried out in ELAN for Windows, versions 4.9.1 – 5.9.

[96] The XML files were first transformed into plain text and were cleaned up, then they were imported into PRAAT to check whether the word segmentation is not too off. Afterwards, the transcriptions were imported to ELAN as PRAAT textgrid files.

English subcorpus							
session ID	duration	speakers present	speakers analyzed	unique speakers	f	m	meeting agenda
EN2002c	0.48.33	3	2	2	0	2	MSc students discuss a group project design
EN2003a	0:37:18	3	2	2	0	2	Three linguistics students plan a postgraduate workshop.
EN2004a	0:57:26	4	2	2	2	0	Four colleagues involved with the AMI project's transcription effort discuss the procedure for checking transcripts.
EN2009b	0:41:06	3	3	3	1	2	Researchers discuss how to get data out of a joint eye tracking system for analysis. The participants include (…) a computer programmer, a senior psycholinguist, and a data processing specialist.
total:	*3:04:23*	*13*	*9*	*9*	*3*	*6*	

Table 4: *English subcorpus – metadata*

Category "unique speakers" refers to the speaker-type frequency, i.e. the number of speakers that have not appeared in other sessions.

The transcription of the Czech subcorpus was performed manually according to the guidelines developed for the DIALOG corpus (Kaderka and Svobodová, 2006), that basically follow the standards of CA. As a whole, the transcription was not synchronized with the other annotations in ELAN at the level of conversational turns, only the segments corresponding to the annotated multimodal constructs (i.e. predicates and immediate left and right context) were captured in a dedicated tier.

5.2.2 Annotated phenomena

To capture the complexity of the mechanism of gesture-speech integration underlying the multimodal expression of eventuality, one cannot simply be content with correlating the gestural boundedness and aspectual types. Therefore, instead of focusing

Czech subcorpus							
session ID	duration	speakers present	speakers analyzed	unique speakers	f	m	meeting agenda
CS021216	0:46:42	3	3	3	3	0	A faculty coordinator and two graduate collaborators discuss evaluation of high school teacher training at the Faculty.
CS020317	0:42:35	5	4	3	1	2	Initial project group meeting of five linguists. The participants discuss methodological issues of spoken corpora.
CS290617	0:52:30	3	3	2	0	2	Faculty management meeting. A vice-dean and two academic administrators discuss a study programm innovation project.
CS300617	1:02:16	2	2	1	1	0	An associate editor of a linguistics journal gives instructions to a newly appointed executive editor.
total:	*3:24:03*	*13*	*12*	*9*	*5*	*4*	

Table 5: *Czech subcorpus – metadata*

only on the aspectual types of gestures, a number of other variables were taken into account that were related to the predicate and may be related to variation in gestural form.

Gesture

Only the *stroke* phases of gestures that co-occurred with predicates were coded. The association between the stroke and the predicate (typically with the verb) was identified primarily on the basis of the prosodic emphasis. If, for instance, the onset of a gesture stroke coincided with a sentence-final verb, but the prosodic emphasis was in the immediately following word, which belonged to the subject NP of the following sentence, the gesture was not attributed to the predicate (see the discussion of alignment of stroke apices and prosodic peaks in Section 2.2).

The annotation category *outer boundedness* was binary: the strokes with a visually

distinguishable ending, either by a stop followed by a hold, or an abrupt change of velocity were coded as *ended gestures* (*e*), the strokes with a lack thereof were coded as *continuous* (*c*).

The annotation category *gesture complexity* was two-fold: it was based on the distinction between *simple* and *complex* gesture forms: simple forms were gestures characterised by a singular stroke without marked movement modulation, complex forms were characterised by the presence of multiple units, either repeated strokes within a complex gesture phrase, or visually discernible movement subphases. This category was also based on the gesture boundedness as defined above: if the internal units involved an abrupt halt of movement, they were attributed the (*e*) value, if the internal units involved a smooth transition (e.g. individual circles of a repeated circular gesture), they were coded as continuous (*c*). The six possible combinations of outer boundedness and complexity values constituted the six gesture types presented in Table 6.

simple forms	**notation**	**code**
ended simplex	–|	e
continuous simplex	–	c
complex forms		
ended with multiple continuous units	~~~~|	ce
ended with multiple ended units	^^|	ee
continuous with multiple continuous units	~~~~–	cc
continuous with multiple ended units	^^–	ec

Table 6: *Six values of the gesture annotation*

For each speaker, separate layers were used for the right and left hand. To every stroke, a gesture boundedness value was then attributed from an inventory (controlled vocabulary in ELAN) of six gesture types. There was no instance of two-handed gestures where gesture boundedness values would be different for individual hands.

Predicates

The referent of the gestural representation often cannot be easily (or reliably) mapped onto a single lexical unit (McNeill, 2005). In many cases, searching for an isolated "lexical affiliate" (Schegloff, 1985) of a gesture or gesture phrase could lead to unwanted reductionism. Therefore, in the context of this study, the *predicate* represents the basic analytic unit, limiting the scope of the potential multimodal construction to the predicate VP. Apart from the aspectual type of the verb (or deverbative N or Adj), other predicate-related categories were taken into account (as they would have been taken anyway in order to interpret the aspectual contour construal of the verbs) as individual categories.

Verbs

In the constructions with auxiliary verbs (such as light verb constructions or constructions with modal verbs), the auxiliary verbs were not annotated unless there was a clear link (via prosodic marking) with the gesture. The gesture accompanied nominal parts of copular constructions were coded if they were deverbative. The verbs and the deverbative N and Adj were attributed an aspectual type and a subtype value and an aspect value.

To operationalize the categories from the modified Croft's model (see Table 2 in the previous chapter), aspectual types were annotated at three levels:

(i) general aspectual type (i.e. the four Vendler classes: *achievements* (ACH), *activities* (ACT), *accomplishments* (ACC) and *states* (STATE));

(ii) aspectual subtype *incrementality* (+/-), applicable only to ACT and ACC;

(iii) aspectual subtype *directedness* (+/-), applicable to ACH and ACH (all ACC were coded as directed).

Aspect was treated as a language-specific category – Czech *perfective* (PFV) and *imperfective* (IPFV), and English *simple* (SIMP), *perfect* (PRF) and *progressive* (PRG).

The annotation scheme took into account the fact that the aspectual types in both languages and aspect in Czech are subject to alternate construals, and that not all contextual cues necessary for assessing the appropriate construal were always available to the coders. Therefore, the coders, when unsure, could give two alternatives and the final decision was made after discussion. Cases where there was no reliable way to attribute a particular aspectual type or aspect even after discussion were discarded from the further analysis.

Modification

It is often the case that the modification of the verb (e.g. by temporal adverbs) shapes the aspectual contour of the event. The question here is to what extent and how gestures interact with the aspectual types with and without adverbial modification; there is a possibility that it is in fact modification that attracts certain gestural patterns.

Adverbial modification of the verbs was coded in terms of *bounding potential* of the modifier. This annotation category had two values: *bounding* and *non-bounding* modifiers. Under the bounding category fall the modifiers of specification (*just*, *exactly*, *explicitly*) and intensification (*very much*, *entirely*), while the vague modifiers such as *normally*, *generally* or *in some way* were coded as non-bounding. Intensification, strictly speaking, is not a matter of BOUNDING, but it is also assumed to be potentially associated with the qualities of simple ended gestures, that tend to be visually salient and, by definition they are accentuated.

[96] See the note on biaspectual verbs below

Complement countability and number

Gestural boundedness does not necessarily have to be associated with the semantics of the verb *per se* but may be related to the semantics of its syntactic complements. Even when the event as a whole is not construed as bounded, some of the entities that take part in the event can be. According to, among others, Langacker (1987b), the perfectivity vs. imperfectivity distinction in verbs is analogous to the between countable and uncountable nouns as both of these concepts are rooted in the same cognitive operation of BOUNDING. In a similar vein, the distinction between a simple (singular) and a complex (repeated) gesture can be related to the number of the complement – a singular gesture stroke representing a singular object.

Complement determinacy

Another area associated with the linguistic encoding of BOUNDEDNESS is a vaguely delimited domain of *determinacy* of the NP – a bounded construal of an object or abstract entity may be signalled by definite articles, demonstrative pronouns or numerals (Langacker, 1987a).

In this study, complement determinacy is treated as a two-tailed category, distinguishing between the *determinate* complements (with demonstrative determiners or with specific reference, e.g. *I finish that page*; *Tohle nikdy neměli*. ('They've never had that.') and *under-determinate* complements (with indefinite pronouns or non-specific reference, e.g. *Something we've moved over*; *Ptali bychom se na různý aspekty*. ('We would ask about various aspects.').

The neutral value (coded as *indeterminate*) comprises mostly complements without determiners. The English definite articles were not automatically annotated as determinate and indefinite as under-determined – thus cases such as *check the dictionary* were treated as indeterminate, while *we're just looking for the two words* was coded as determinate.

Deixis

Outside the domain of boundedness, *deictic* expressions, in general, are of interest here as well, since, as has been discussed in detail in Chapter 2, iconic representation and referentiality in gesture is often intricately conflated. It was also shown that referential gestures sometimes have similar formal properties to bounded gestures (burst movement culminating in an apical hold, see Cooperrider, 2011). In the following analysis, the potentially conflated gestural indexicality and event BOUNDARY construal is treated separately. Deixis was annotated simply as presence or absence of deictically referential expressions within the predicate.

Negation

Finally, there is a number of ways in which *negation* might be related to the occurence of *e*-gestures. For instance, in terms of information structure, a negated predicate might be contrastivelly focused (against a positive presupossition) – in that case, a presence of gesture may be related to marking of the constrastive emphasis. Verbal negation was shown to be co-expressed gesturally in several ways (Kendon, 2004; Harrison, 2014), some of which involve an abrupt halt in the movement (Calbris, 2003), or are characterized by a holding phase (Bressem and Müller, 2017).

Table 7 provides an overview of the annotated categories and their values.

variable	**code**	**values**	**code**
aspectual type	`vendler`	achievement, accomplishment, activity, state	`ach, acc, act, state`
incrementality	`inc`	incremental, non-incremental	`inc, noninc`
directedness	`dir`	directed, undirected	`dir, undir`
gesture boundedness	`gest_out`	ended, continuous	`e, c`
gesture complexity	`gest_comp`	simple, complex	`simple, complex`
complement number	`obj_num`	singular, plural	`sg, pl`
complement countability	`obj_count`	countable, uncountable	`count,uncount`
complement determinacy	`obj_det`	underdeterminate, determinate, indeterminate	`undet, det, indet`
modifier boundedness	`modif_bd`	bounding, non-bounding	`bd, ubd`
negation	`negation`	yes, no	`yes, no`
aspect	`aspect`	English: progressive, perfect, simple	`prg, prf, simp`
		Czech: imperfective, perfective	`ipfv, pfv`

Table 7: *Overview of the variables*

5.2.3 Inter-annotator agreement

Degree of agreement among annotators is a measure of coding reliability. Whenever annotation involves coding of categories based on subjective/introspective/intuition-based criteria, it needs to be performed independently by at least two coders, in order to be able to quantitatively assess the degree of agreement between their initial coding, before unanimity is achieved. The higher the degree of initial agreement, the more likely is the coding scheme to yield similar results when replicated. The decision upon the final coding of the instances of disagreement should ideally be also based on quantitative measures, but it requires more than two coders to reach a majority. In this study, a segment of the coding was performed by two independent coders in order to assess the degree of agreement. Unanimity was reached through a discussion over the problematic points in the entire dataset.

A standard way of quantitative assessment of inter-rater agreements is calculation of one of the agreement coefficients rather than calculation of percentual proportion of agreement. The disadvantage of simple percentual proportion is that it is prone to be skewed towards higher reliability estimation when the number of coded instances (observations) gets higher – as the probability of reaching an agreement by chance increases. Also, the risk of unaccounted chance agreement is related to the complexity of the coding scheme: with more levels within one coding category, the probability of chance agreement decreases. Therefore, agreement coefficients that take both the number of observations as well as complexity of the coding scheme into account should be preferred.

For measuring the agreement between two annotators coding nominal variables, the most widely used statistic is Cohen's κcoefficient (Cohen, 1960). Here, individual κscores were calculated in R (R Core Team, 2020) with the help of `kappa2()` function from *irr* package (Gamer et al., 2012).

There is not a single accepted way of interpreting κscores. I will follow the (arbitrary, but widely adhered to) division of the levels of agreement according to κ values proposed by Landis and Koch (1977, p. 164–165):

κ<	0.00	poor agreement
κ=	0.00–0.20	slight agreement
κ=	0.21–0.40	fair agreement
κ=	0.41–0.60	moderate agreement
κ=	0.61–0.80	substantial agreement
κ=	0.81–1.00	almost perfect agreement

For the present study, the threshold value is set to κ= 0.50. Should the agreement not meet this limit, the annotation operationalization for the given category will have to be revised and the annotation performed again.

Gestures

Out of 575 gesture phrases, 101 were randomly selected (i.e. 18% of the entire dataset, 54 gesture phrases from the English subcorpus, 47 from the Czech subcorpus) to be annotated independently by two coders. The degree of agreement was substantial with κ= 0.79 and raw agreement in 75.25% cases. If we take into account also the partial agreement, i.e. instances when the two coders were in accord with respect to only one annotation parameter (outer or inner boundary, i.e. agreement 0.50) the raw agreement rate rises to 82.18%. When it comes to distinguishing *simple* and *complex* gesture forms, the coders reached an agreement in 89.11% of cases. Such a level of agreement suggests that the parameters of the annotation scheme were well operationalized – crucially, the formal features selected for annotation were indeed distinguishable by human coders, even despite the relatively limited video quality of the AMI corpus (raw agreement 72.22%).

Eventuality types

In the case of eventuality types, 110 observations were randomly selected from the dataset (19.20%, balanced for the two languages). The degree of inter-annotator agreement was lower than in the case of gestures, but still acceptable: Cohen's $\kappa = 0.55$ (i. e. moderate agreement), with raw agreement rate at 64.49%. A relatively weaker agreement was to be expected: aspectuality types are rather elusive semantic constructs as the speaker may profile the same event as textscach or ACT, and the actual construal depends on a number of contextual cues, including gestures. Gestures, as the dependent variable in the present study, could not be taken into account when coding eventuality. The coders reviewed the points of disagreement, and the cases that could not have been resolved without gestural input were set aside.

The second problem to consider here is the fact that how eventuality is actually construed may be affected, among other contextual cues that were not captured in the corpus annotation, by gestures. From the methodological point of view, gestures cannot be taken into account when coding eventuality types as they are the dependent variable in the context of the present study.

5.3 Analysis

The relationship between the variables was explored by means of the *conditional inference trees* and *random forests* (introduced in the corpus linguistics context by Tagliamonte and Baayen, 2012; cf. also recent critical discussion by Gries, 2019) – a non-parametric alternative to logistic regression models suitable for the datasets violating the regression assumptions, such as normal distribution, lack of correlation between variables or homogeneity of residual variance (*heteroscedasticity*).

As the dataset in this study is characterised by several inter-correlated variables, nested variables (*incrementality* and *directedness*) and a relative sparsity of observations per category – the conditional inference tree and random forest models represent an ideal way of multifactorial analysis. Although superior to logistic regression in many respects, this method has a potential pitfall in its tendency for overfitting, which must be always born in mind when drawing conclusion from the smallish datasets as the one we are dealing with here.

First, using `cforest()` function from the *party* package (Strobl et al., 2008) in R, random forests of conditional inference trees for the two response variables – *gesture boundedness* and *gesture complexity* were generated separately for English and Czech. After generating a random forest, it was possible to assess the conditional importance of the individual predictors (using `varimp()` function of the *party* package): *Vendler classes*, *incrementality*, *directedness*, *aspect*, *complement number*, *complement determinacy*,

[96] In this context, conditional importance means that the variables were tested while taking into account their correlation.

modifier boundedness, *negation* and *deixis*. *Complement countability* was not included in the final analysis, as almost all instances in the dataset were countable.

To interpret the effects of the individual predictors and their interactions, another model based was then fit, based on a solitary conditional inference tree (henceforth "tree") for each of the four combinations of language and response, using `ctree()` function from the *partykit* package (Hothorn et al., 2006). Although a less accurate representation of the data than the random forest models, the trees allow for, with a grain of salt, teasing apart the effects and better understand the relationship between the predictors.

5.3.1 Data exploration

In total, 575 multimodal units (multimodal constructs) were annotated: 332 in the English subcorpus, 243 in the Czech subcorpus. Nineteen cases were discarded from the final analysis due to coding issues – these cases were instances where gestural input was necessary for attributing an aspectual type (i.e., they represented instances of *profiling-for-discrimination*).

The final dataset thus consisted of 556 observations, 328 from the English sample, 228 from the Czech sample.

The number of the analysed multimodal constructs (see Figure 19) varied considerably across speakers. In the English subcorpus, the number of verbs accompanying gestures produced by individual speakers ranged between 11 and 55, while the Czech subjects produced between 13–84. As is evident from the below visualization of the distribution, the upper limit value in the Czech sample deviates from the overall tendency which is here best represented by the median value (36 for English and 18 for Czech). The fact that an outlier speaker produced almost 35% of gestures in the Czech subsample must not be disregarded as it may affect the results of the subsequent quantitative analysis.

Verbs

Let us first consider the linguistic component of the multimodal constructs. Table 8 shows type-token ratios for the verbs accompanied by gestures in the two subcorpora, disregarding the auxiliary verbs. The type-token ratio is considerably higher in the Czech subcorpus (0.72 vs. 0.52). However, one must take into account that in Czech, the aspectual variants are considered independent lexemes, which may have caused the

[96] *Partykit* evolved from the older *party* package. While both packages feature tools for random forest as well as conditional inference tree analyses, *Partykit* is superior for modelling trees, but does not allow for computing conditional importance for correlated variables in random forests.

[96] More detail on these cases will be given in the discussion (Section 5.5).

[96] For the purpose of type-token ratio calculation and semantic analysis, the deverbative nouns and adjectives contained in the sample are substituted by the respective base verbs.

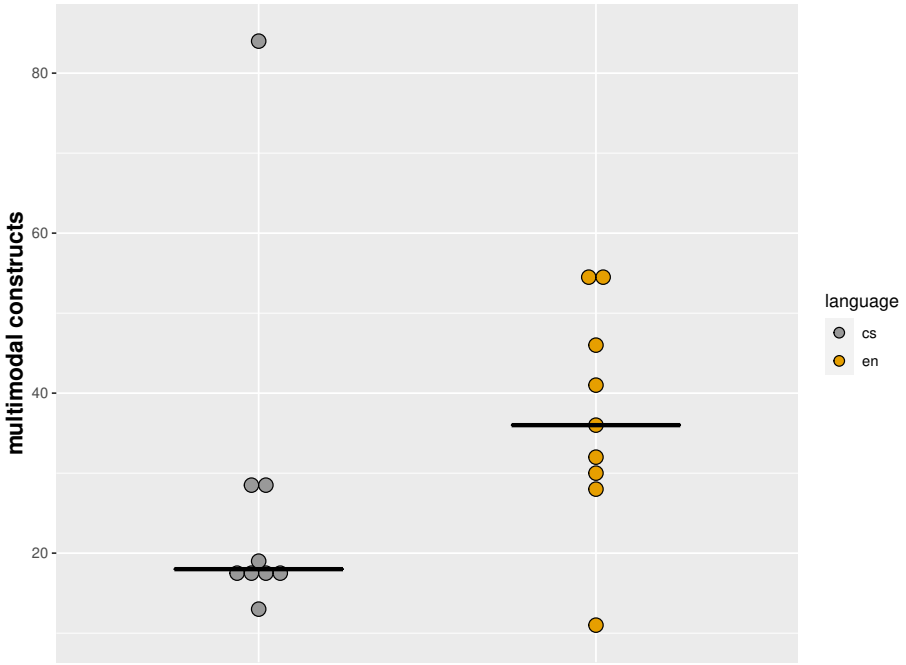


Figure 19: *Number of analysed multimodal constructs by speakers in both subcorpora*

inflation of verb-type density in the Czech sample. The type-token ratio in the Czech subcorpus is higher because Czech is a morphologically much richer language than English and thus, for instance, the different aspectual variants are treated as separate types. When the aspectual variants are counted as realizations (tokens) of a single type, the difference between English and Czech decreases (0.66 vs. 0.52).

	English	**Czech**	**Czech (aspectual variants disregarded)**
Verbs-tokens	332	243	243
Verbs-types	169	175	160
Type/token ratio	0.52	0.72	0.66

Table 8: *Verb types vs. tokens in the two subcorpora*

Tables 9 and 10 capture the most frequent lemmata in the two subcorpora.

Both samples are characterized by the most frequents verbs belonging to the three semantic groups: *verba cogitandi* (mental actions/states), *verba dicendi* (communication) and motion verbs, expressing either actual or fictive motion in metaphoric extensions. Such a distribution of semantic classes appears to be expectable as the sessions included in both subcorpora share the same genre and setting.

lemma	translation	frequency	aspect
vědět	'know'	10	IPFV
říci	'say'	5	PFV
dostat	'get'	4	PFV
jít	'go'	4	IPFV
mít	'have'	4	IPFV
ptát se	'ask'	4	IPFV
učit	'learn'	4	IPFV
dávat	'give'	3	IPFV
fungovat	'work' (function)	3	IPFV
plánovat	'plan'	3	IPFV

Table 9: *Ten most frequent verbs: Czech subcorpus*

lemma	frequency
say	11
create	8
do + OBJ	8
run	8
use	8
look at	7
look for	7
go through	6
know	6
add	5

Table 10: *Ten most frequent verbs: English subcorpus*

Eventuality types

The mosaic plot (Figure 20, see also Table 11 below) provides a general picture of the distribution of aspectual types in the two samples, represented by the four Vendler classes. An association test revealed a significant difference between the two samples in the distribution of aspectual types ($\chi^2_{(3)} = 23.03, p < 0.001$). The shading of the tiles in the mosaic plot represents the difference between expected and observed frequencies (Pearson residuals), red tiles correspond to Pearson residuals ≤ 1.96, blue to ≥ 1.96. ACH are the most frequent aspectual type in the English sample, whereas in Czech, the most frequent type is ACT.

Given the nature of the data, collected in an uncontrolled, naturalistic setting, a relatively high degree of inter-speaker variation is inevitable. In the context of this study, an important question is whether the degree of variance in the two subcorpora is comparable. The diagram 21 and Table 12 show the proportions of Vendler types averaged by subjects, together with the variance (and underlying distribution). We can see a greater amount of dispersion in the Czech dataset. However, when we inspect the individual proportion of aspectual types per subject (Figure 22), an outlier speaker can be identified (3_m_cs) deviating in the frequency of ACH. Disregarding this particular

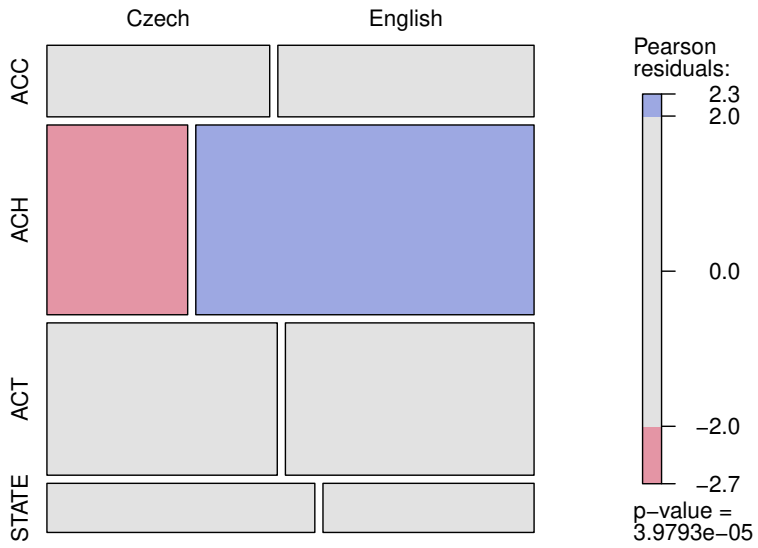


Figure 20: *Distribution of aspectual types (Vendler classes) types in the two subcorpora*

	aspectual types	n	proportion	Pearson residual
Czech	ACC	40	0.18	0.80
	ACH	67	0.29	-2.74***
	ACT	88	0.39	1.50
	STATE	33	0.14	1.79
English	ACC	46	0.14	-0.66
	ACH	161	0.49	2.28***
	ACT	95	0.29	-1.25
	STATE	26	0.08	-1.49

Table 11: *Distribution of aspectual types (Vendler classes) types in the two subcorpora* (*** = significant values of Pearson residuals, shading of the tiles in the mosaic plot represents the magnitude and direction of Pearson residuals (+/-1.96 = significant residual value).

speaker, whose overall contribution is rather small (13 gestures), the above-mentioned tendencies of ACH and ACT to be the most frequent gesture-accompanied aspectual types in the English and Czech subcorpus, respectively, are manifested across speaker in a relatively uniform manner.

Gestures

All six possible combinations of the annotated gestural features were represented in the data. Relative frequencies of the gesture types were comparable across the two

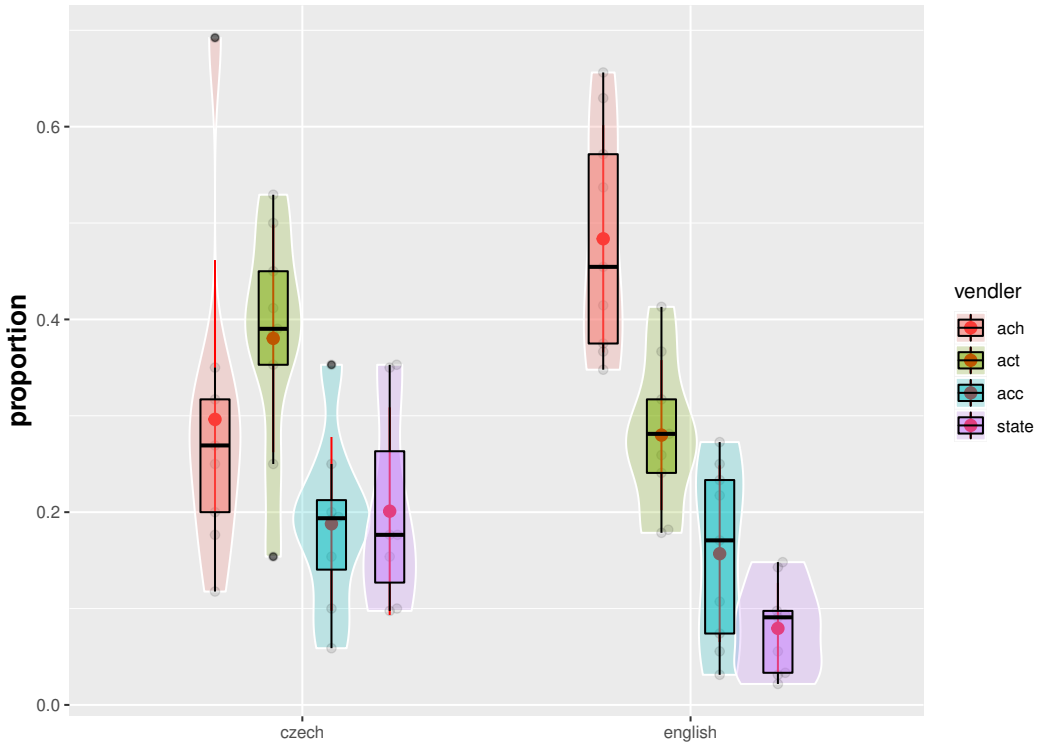


Figure 21: *Inter-speaker variation – aspectual types (Vendler classes) types of the analysed multimodal constructs*

(Red dots = mean proportions (averaged by subjects), vertical lines = SD, boxes = interquartile range, horizontal bars = median, grey dots = subjects)

	aspectual types	mean proportion by subject	SD
Czech	ACC	0.17	0.11
	ACH	0.30	0.17
	ACT	0.38	0.12
	STATE	0.16	0.13
English	ACC	0.16	0.09
	ACH	0.48	0.19
	ACT	0.28	0.08
	STATE	0.08	0.05

Table 12: *Inter-speaker variation – aspectual types (Vendler classes) types of the analysed multimodal constructs*

SD = standard deviation

languages (Table 13). Calculation of the Pearson residuals, that account for the magnitude of difference between observed and expected frequencies revealed no significant crosslinguistic difference ($\chi^2_{(5)} = 10.86$, $p = 0.054$).

The most frequent gesture type in both languages was ended simplex (*e*, 59% and 80% of all co-verb gestures in the English and Czech subcorpus, respectively), followed by *cc* (21% in the Czech subcorpus and 12% in the English). Except for *ce* gestures in

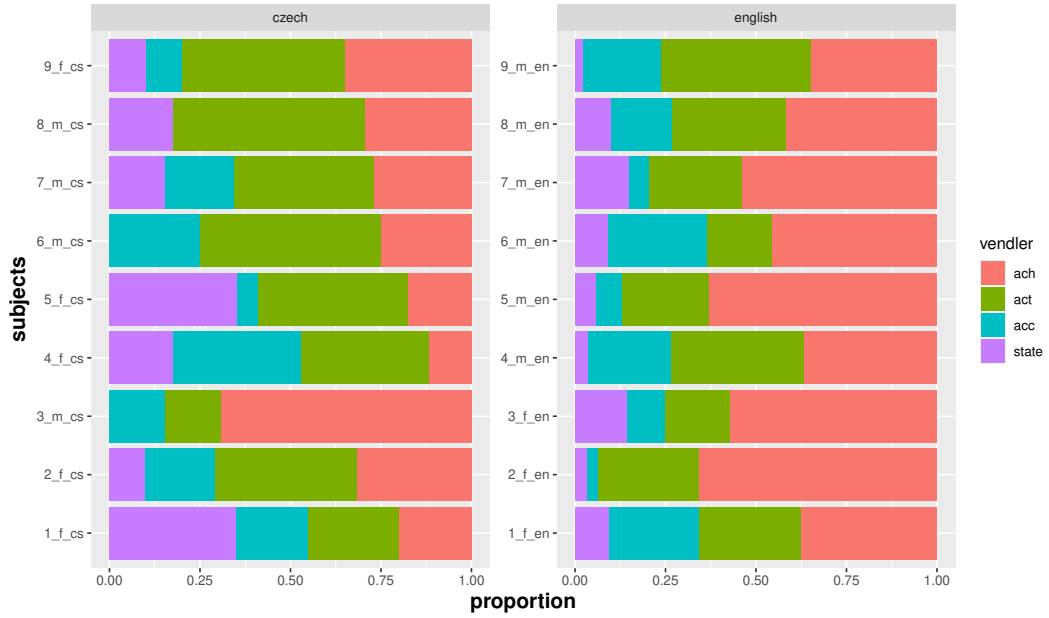


Figure 22: *Proportions of aspectual types – individual subjects*

	gesture	n	Pearson residual	relative frequency
Czech	e	121	0.91	0.50
	ee	9	-0.31	0.04
	ec	16	1.79	0.07
	c	23	-1.16	0.09
	cc	51	-0.71	0.21
	ce	23	0.55	0.09
English	e	197	-0.78	0.59
	ee	18	0.26	0.05
	ec	17	-1.53	0.05
	c	22	0.99	0.07
	cc	43	0.61	0.13
	ce	35	-0.47	0.11

Table 13: *Distribution of gesture types in the two subcorpora*

the English sample, the proportion of any of the remaining types did not surpass 10%.

Ended gestures (types *e*, *ee*, and *ec*) predominate in both subcorpora (in the English sample, they make 69% of all co-verb gestures, in the Czech sample 60%), and unlike the continuous group, they typically occur in the simplex form. This is not at all surprising – ended gestures highlight primarily a single point – the event's end – and are often characterized by a sudden acceleration of movement and thus afford more for a one-off realization.

Interestingly, the complex gestures did not exhibit a tendency to occur in consistent combinations of outer and inner boundary features (*cc*, *ee*). In fact, the inconsistent combinations (*ce*, *ec*) had even higher (Czech) or the same (English) frequencies than the consistent ones, suggesting that the dual view of gesture boundedness is justified

and may provide a better insight to the embodied qualities of event construals.

Figure 23 (see also Table 14) visualizes the between-subject variance in the proportion of the individual gesture types and Figure 24 displays the proportions in subjects.

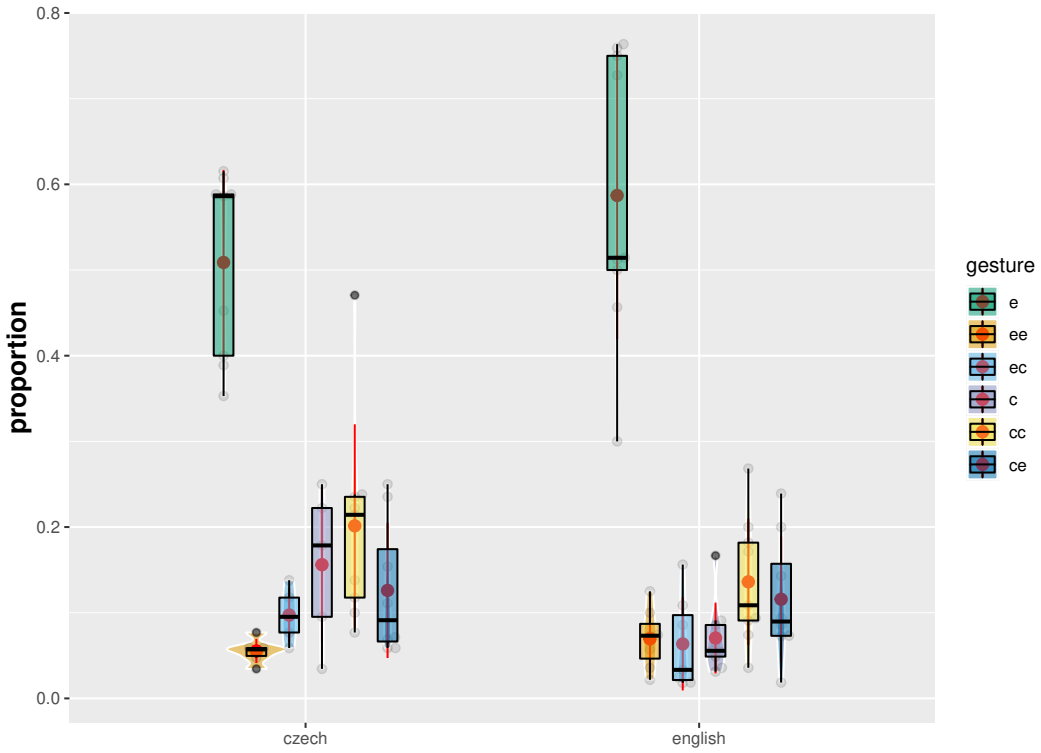


Figure 23: *Inter-speaker variation – gesture types*
(Red dots = mean proportions (averaged by subjects), vertical lines = SD, boxes = interquartile range, horizontal bars = median, grey dots = subjects)

	gesture	by-subject mean prop.	SD
Czech	e	0.51	0.11
	ee	0.06	0.01
	ec	0.10	0.03
	c	0.16	0.09
	cc	0.20	0.12
	ce	0.13	0.08
English	e	0.59	0.17
	ee	0.07	0.04
	ec	0.06	0.05
	c	0.07	0.04
	cc	0.14	0.07
	ce	0.12	0.07

Table 14: *Distribution of gesture types in the two subcorpora - by-subject mean proportions*
SD = standard deviation

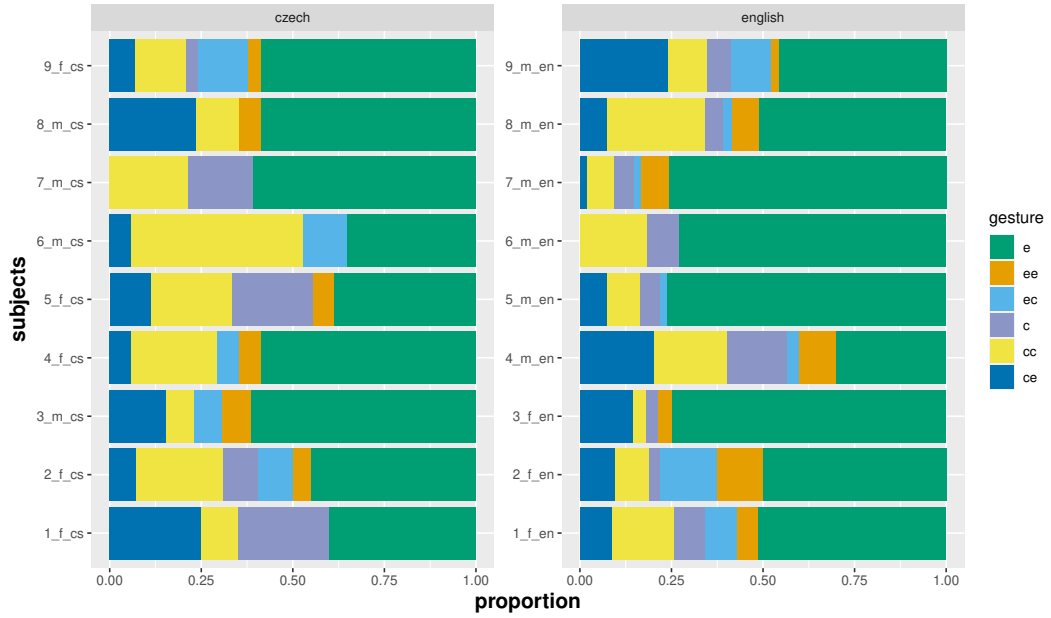


Figure 24: *Proportions of gesture types – individual subjects*

The predominance of *e*-gestures is evident across speakers, as well the relatively low overall degree of variance. Given the limited number of subjects and particularly the imbalance in the number observations per subject, a relative uniformity across subjects is an important prerequisite for the subsequent statistical analysis (especially when the inter-speaker variance cannot be controlled, e.g. via the inclusion of random effects in regression models). All in all, the level of inter-speaker variance, concerning both aspectual types and gestures, does not represent a serious issue. Nevertheless, I will remain careful not to draw too general conclusions upon the present dataset, as there is room for improvement regarding representativeness of the sample.

Outer boundedness, complexity and Vendler classes

One of the aims of this study is to investigate the co-occurrence of gesture form and semantic features at a level of granularity beyond the Vendler classes. Prior to proceeding with the multifactorial analysis that will enable us to do so, a first-glance picture can be made from reviewing how the four Vendler classes combine with the gestural types in the two samples. Table 15 gives the relative frequencies of the particular combinations.

Focusing on the two most frequent gesture types, we can see that in English, more the majority of *e*-gestures is associated with ACH, whereas in Czech, the proportion of *e* + ACH combinations is lower, closer to the proportion of *e* + ACT. The second most frequent type, *cc*, has practically the same distribution across the Vendler classes in the two languages. The Cochran-Mantel-Haenszel test (Cochran, 1954), an elaboration of the χ^2 test can be used to assess whether the observed frequencies of the

		Vendler class			
	gesture type	**ACH**	**ACT**	**ACC**	**STATE**
Czech	e	0.40	0.34	0.16	0.10
	ee	0.33	0.44	0.22	0.00
	ec	0.20	0.40	0.40	0.00
	c	0.23	0.27	0.14	0.36
	cc	0.22	0.43	0.15	0.20
	ce	0.00	0.59	0.18	0.23
English	e	0.61	0.20	0.12	0.07
	ee	0.50	0.28	0.17	0.06
	ec	0.19	0.69	0.13	0.00
	c	0.41	0.32	0.14	0.14
	cc	0.28	0.47	0.14	0.12
	ce	0.29	0.37	0.23	0.11

Table 15: *Relative frequencies of gesture type – Vendler class combinations*

combinations of the two categorical variables (in this case, gesture and aspectual types) differ across strata (languages). The difference between English and Czech is significant ($M^2{}_{(15)}$= 64.60, $p < 0.001$). Figure 25 allows for a closer look at what exactly contributes to the significant difference. The mosaic plot visualizes the magnitudes of the differences between expected and observed frequencies (Pearson residuals) compared between the two languages (via separate χ^2 tests for the individual aspectual types).

The *e* + ACH combination has indeed a significantly higher relative frequency in English, whereas the *cc*-type is underrepresented. With ACT, *e*-gestures occurred relatively less in the English sample, cc gestures had higher-than-expected frequency in the Czech data. The *ec*-type, while marginal in Czech, had a significant association with ACT in English. ACC exhibits a comparable distribution of gesture types (*ec*-gestures are overrepresented in Czech but that accounts for only 6 instances). STATE have a significantly stronger association with *c*-types in Czech.

Breaking up the gestures into two clusters according to the outer boundedness (*e*- vs. *c*- types) yields a very similar picture (Table 16 and Figure 26).

		Vendler class			
	boundedness	**ACH**	**ACT**	**ACC**	**STATE**
Czech	e	0.38	0.36	0.19	0.08
	c	0.17	0.43	0.16	0.24
English	e	0.57	0.24	0.13	0.06
	c	0.31	0.40	0.17	0.12

Table 16: *Relative frequencies of gesture boundedness – Vendler class combinations*

The *e*- types are significantly ($M^2{}_{(3)}$= 37.244, $p < 0.001$) associated with ACH in English, while the *c*-types are underrepresented with ACH in both languages. ACT and ACC, classes sharing the feature of *durativity* and the *incrementality* distinction, while still more frequently co-occurring with *e*-gestures, have a greater proportion of *c*-

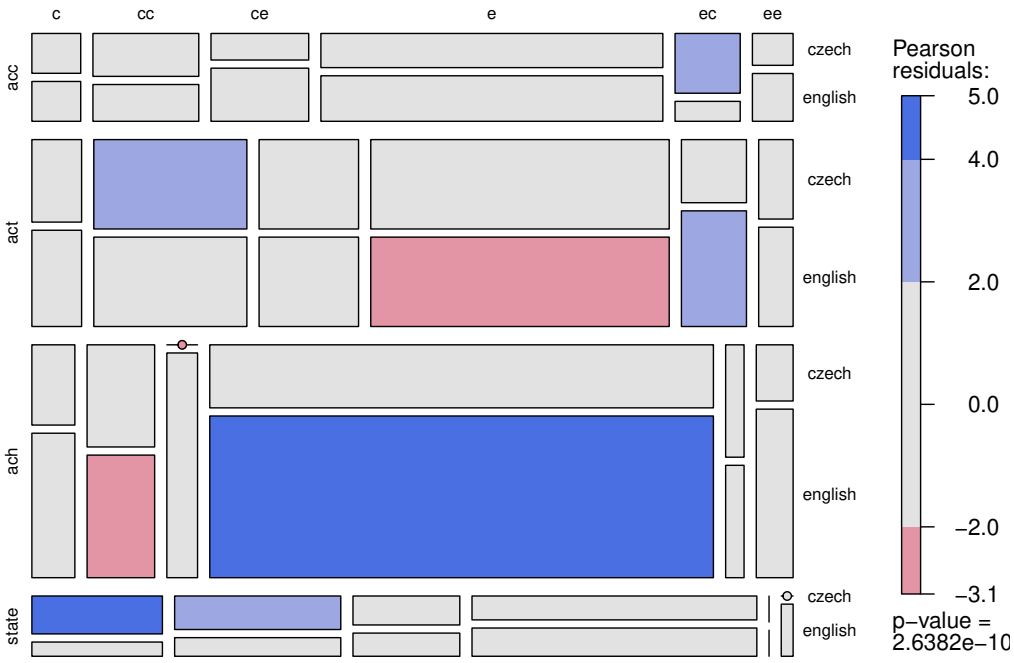


Figure 25: *Mosaic plot – associations between gesture types and Vendler classes in English and Czech samples*

gestures. In Czech, the observed frequency of *c*-gestures was significantly greater than expected. Lumped together, the distribution of outer boundedness types with ACC does not show a significant difference, while with STATE, *c*-gestures are significantly more frequent in Czech.

Finally, let us focus on gesture complexity. The relative frequencies of simple and complex forms with the particular Vendler classes are given in the Table 18 below.

		Vendler class			
	complexity	**ACH**	**ACT**	**ACC**	**STATE**
Czech	simple	0.38	0.33	0.15	0.14
	complex	0.17	0.47	0.21	0.15
English	simple	0.59	0.21	0.13	0.07
	complex	0.30	0.44	0.17	0.09

Table 17: *Relative frequencies of gesture complexity – Vendler class combinations*

Showing a similar pattern (not surprising given the correlation between boundedness and complexity – see below), the mosaic plot (Figure 27) reveals significant associations between complexity and ACH and ACT ($M^2{}_{(3)} = 36.422$, $p < 0.001$). In the former case, simple gestures are significantly overrepresented in English and complex gestures are significantly underrepresented. In both languages, there are twice as many

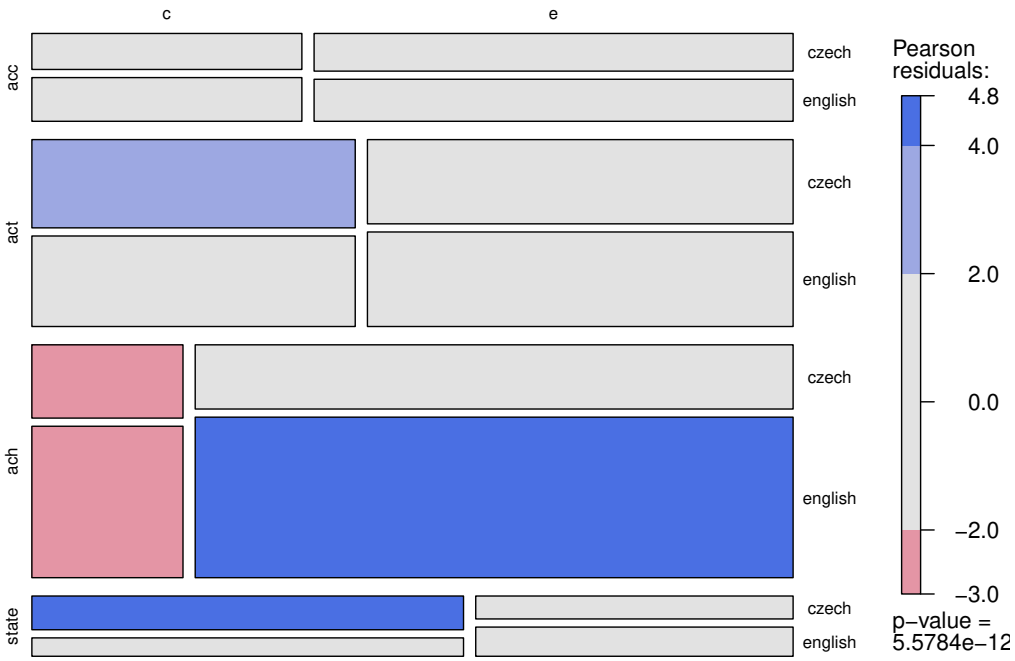


Figure 26: *Mosaic plot – associations between gesture boundedness and Vendler classes in English and Czech samples*

simple gestures with ACH than complex one, in English the majority of all simple gestures (60%) occurred with ACH. An opposite tendency was found with ACT in both languages. The positive association with complex gestures was significant in Czech, in English there was a significant value of a negative Pearson residual in *simple* + ACT combination.

As is evident from the distribution of boundedness-feature combinations (Table 18, the majority of *c*-gestures occurred in complex forms – 77%, compared to only 16% in the case of *e*-gestures. Importantly, this pattern was present in both languages: Table 15 shows that the proportions of complex and simplex forms across in *e*- and *c*-gestures is practically the same in English and Czech samples.

		outer boundedness	
	complexity	**e**	**c**
Czech	simple	0.83	0.24
	complex	0.17	0.76
English	simple	0.85	0.22
	complex	0.15	0.78

Table 18: *Crosstabulation of outer boundedness and complexity*

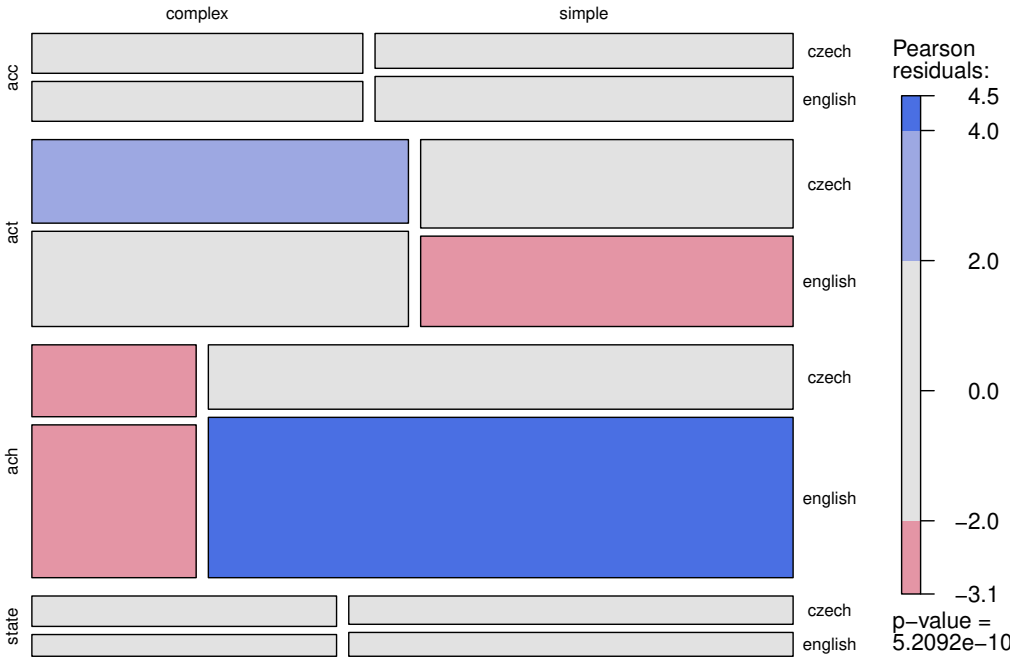


Figure 27: *Mosaic plot – associations between gesture complexity and Vendler classes in English and Czech samples*

5.4 Results

5.4.1 English

The first random forest model (model 1) was fit for outer boundedness and all predictors except complement countability. As for the accuracy of prediction, calculated as the percentage of correctly predicted gesture boundedness values, the model performs well above the chance level (73.78%), the model's index of concordance is $C = 0.77$, which means a fairly good discrimination level. Figure 28 shows the relative conditional importance of the individual predictors, estimated as average loss of accuracy of the model when the variable is removed (*mean accuracy decrease* method, Breiman, 2001). The dashed line separates the relevant variables on the right-hand side of the plot from the variables that do not have an effect on the predictive performance of the model (the division value corresponds to the absolute value of the lowest importance score (Levshina, 2015, p. 298).

[96]The model formula in R syntax is `gest_out ~ vendler + aspect + inc + dir + obj_det + object_num + modif_bd + negation`.

[96]Note however, that while the model predicts the distribution of *e*-gestures accurately, it fails in predicting *c*-gestures, see Table 19.

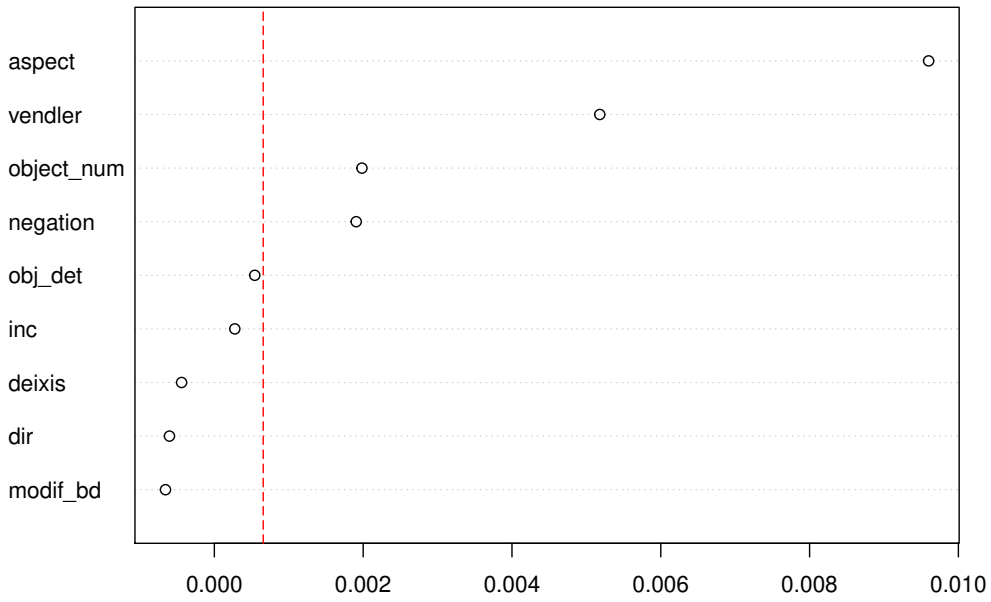


Figure 28: *Predictor conditional importance – model 1*

According to the random forest model, aspect and Vendler classes are the most important predictors, followed by the number of the complement and negation.

Fitting a model based on a single conditional inference tree (with the irrelevant variables removed) sheds some light on how the relevant factors interact. Notwithstanding the worse discrimination performance (accuracy 69.82%), the model (Figure 29) estimates the first significant binary split in the data between ACH on the one hand (80.75% of *e*-gestures) and the remaining Vendler classes on the other (58.68% of *e*-gestures). Aspect comes into play in the non-ACH group, where progressive ACT have a lower proportion of e-gestures (49.09% vs. 63.39%, split 2). Another decrease in e-gestures is marked by either plural or no complement present with ACC, STATE and non-progressive ACT (50.00% vs. 75.86% of *e*-gestures when the complement is singular (split 3)). The final significant split is within the ACH cluster, conditioned by negation – the absence of negation has a stronger association with *e*-gestures than the presence of negation.

For gesture complexity, random forest model (model 2, Figure 30) identifies a clear cluster of important predictors with a prominent role of incrementality and Vendler classes. A tree model with only these two variables included (figure omitted) fails to predict complex gestures (0% accuracy). When we include aspect and deixis

[96] formula: `gest_comp ~ vendler + aspect + inc + dir + obj_det + object_num + modif_bd + negation`

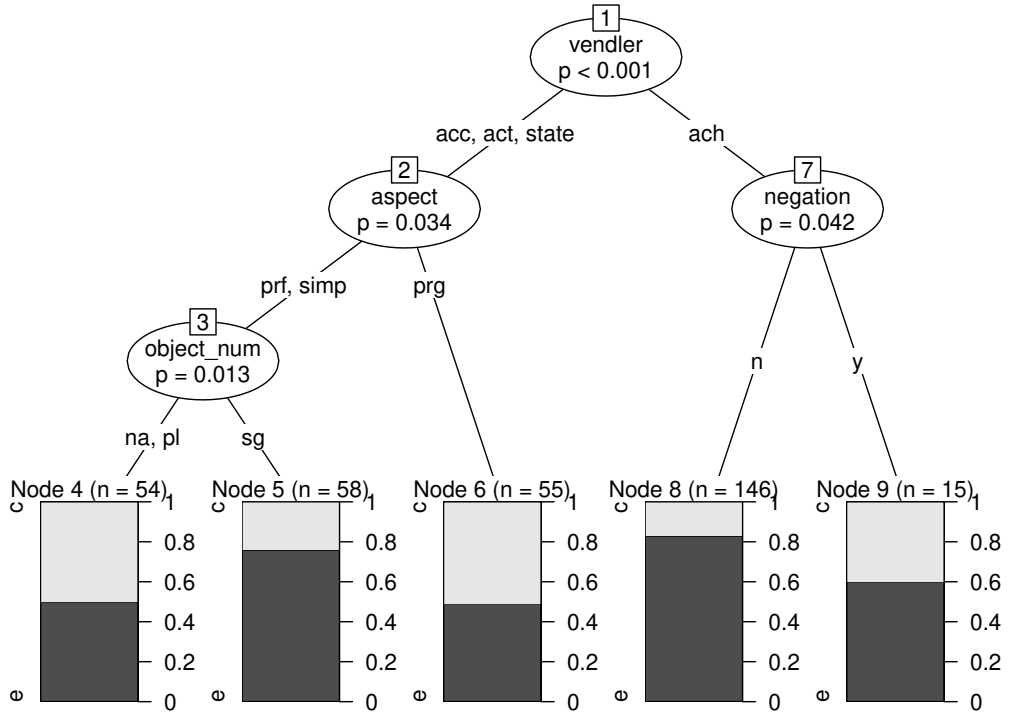


Figure 29: *Tree model 1*

(Figure 31), the performance of the model improves.

Inspecting the data, we can see that there are in fact two concurrent tendencies within the non-achievement cluster: first, the preference of progressive ACT for complex gestures (56.36% complex gestures with PRG vs. 29.67% with other aspects), second, there is a preference of incremental ACC to attract more complex gestures (46.15%) compared to non-incremental ACC (39.39%) (there is only a single instance of an incremental ACT in the English sample).

5.4.2 Czech

Following the same procedure with the Czech sample, I first report the relative importance of the predictors based on the random forest model of gesture *outer boundedness* (model 3, Figure 32).

Incrementality and directedness stand out as the most important predictors of outer boundedness in the Czech data. Vendler classes, aspect, deixis and the number of the complement also had non-zero scores but they were outperformed by the finer-grained aspectuality features.

The tree (Figure 33) brings out the nature of interactions between incrementality and directedness, and allows us to interpret the relevance of Vendler classes and aspect in the Czech dataset. The first split (1) again partitions the Aktionsart classes. In Czech, the distinction between ACH and ACC on the one hand and ACT and STATE on

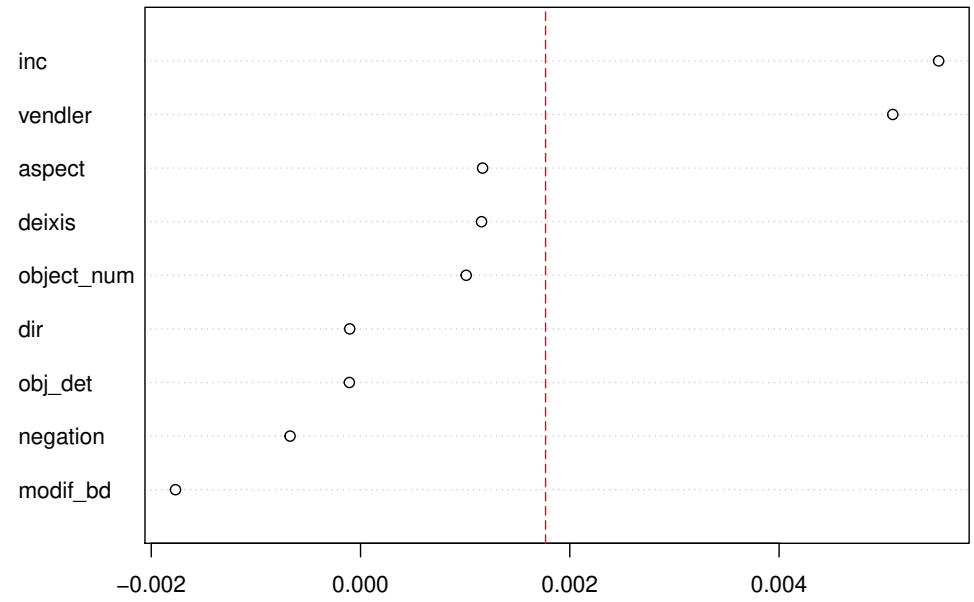


Figure 30: *Predictor conditional importance – model 2*

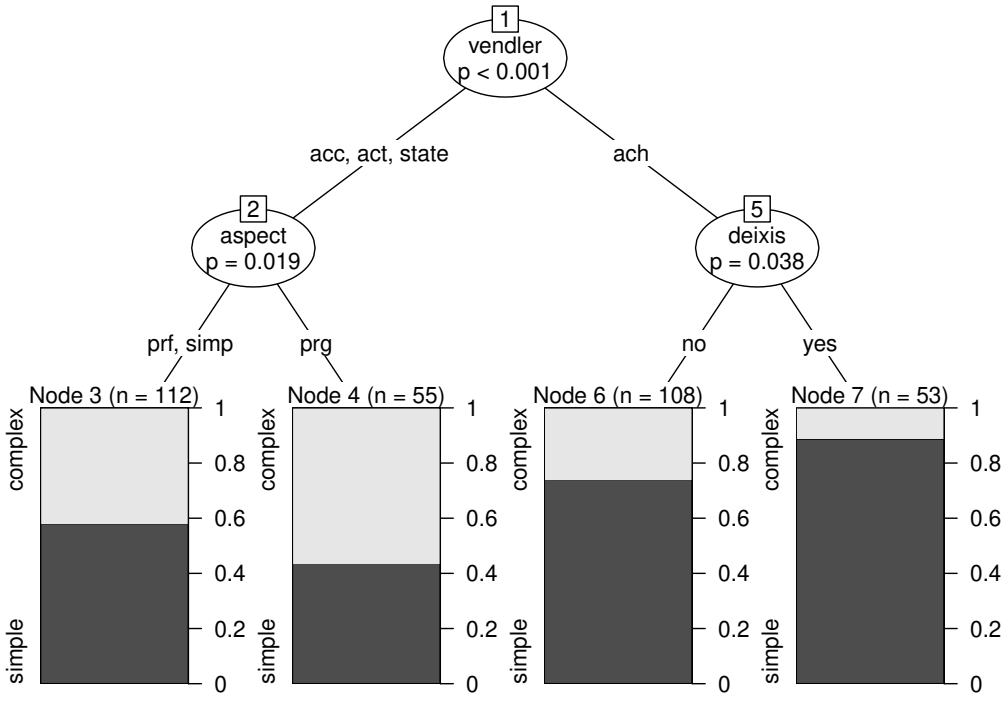


Figure 31: *Tree model 2*

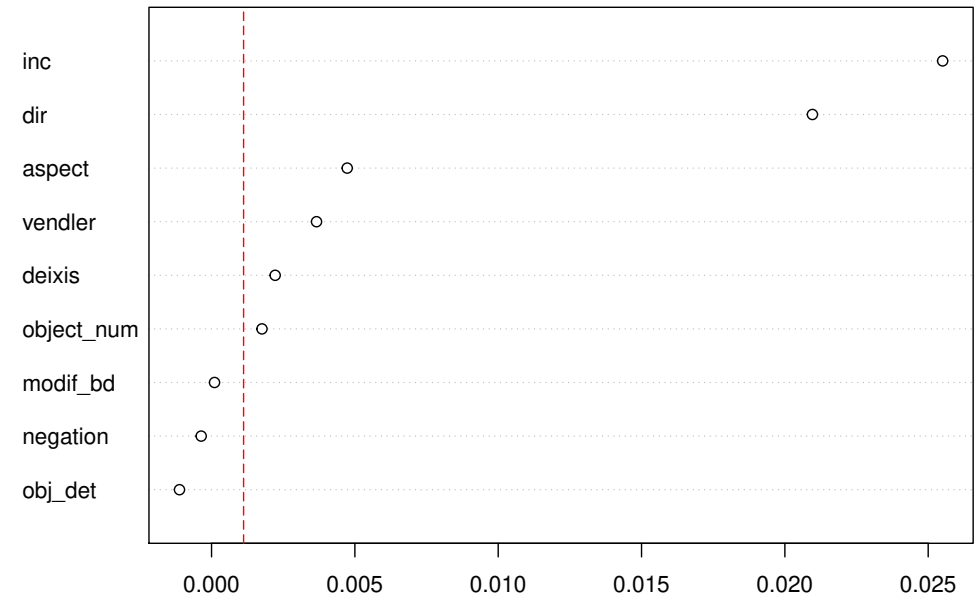


Figure 32: *Predictor conditional importance – model 3*

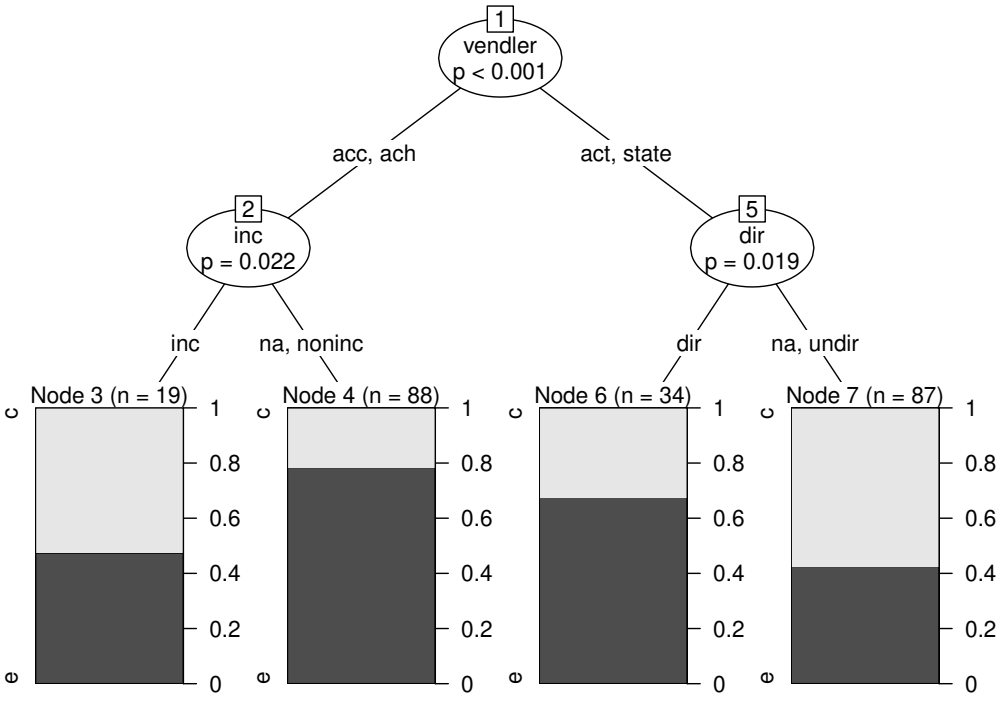


Figure 33: *Tree model 3*

the other mirrors the distinction between PFV and IPFV, respectively. Along this division, the contrast in outer boundedness marking is clearly visible: 72.90% of gestures in the PFV cluster (2) have a marked ending, compared to 49.29% in the IPFV cluster (5). However, it is not about the aspect alone; based on the interaction of incrementality and directedness, the tree model suggests three clusters of event types that tend to co-occur with ended gestures to different degrees:

(i) ACH, non-incremental ACC, directed non-incremental ACT (77.78% accompanied by *e*-gestures)

(ii) incremental ACC, undirected ACT (47.95% accompanied by *e*-gestures)

(iii) directed incremental ACT, STATE (33.33% accompanied by *e*-gestures)

Finally, the random forest model for gesture complexity (model 4, Figure 34) shows that directedness is by far the strongest predictor of gesture complexity in the Czech sample, followed by incrementality and negation.

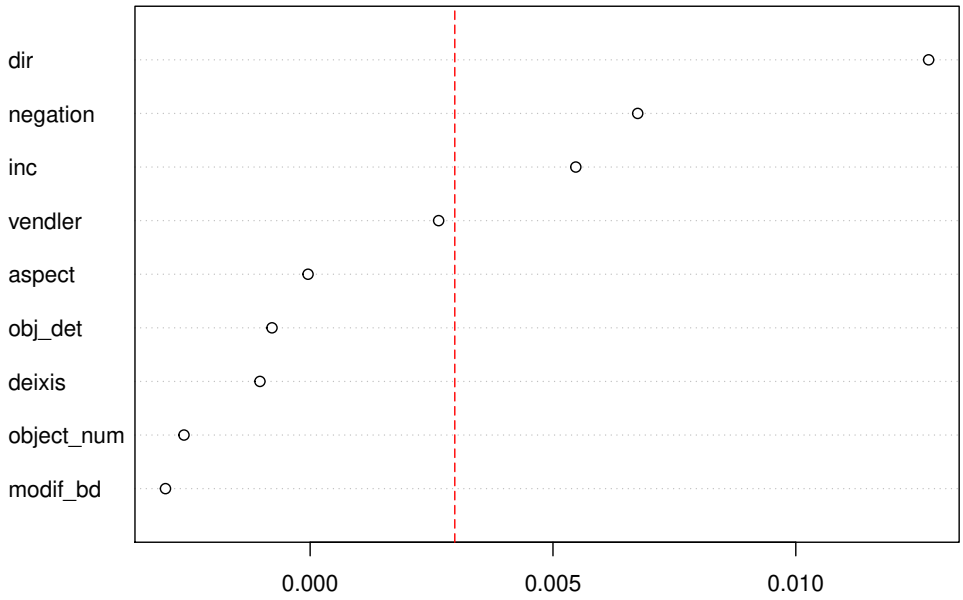


Figure 34: *Predictor conditional importance – model 4*

The tree (figure omitted) is made up of only two terminal nodes with a single significant split between undirected (56.14% of complex gestures) and directed subtypes combined with the remaining types (35.09% of complex gestures).

The random forest model suggests that gesture complexity with directed verbs also depends on incrementality; incremental ACC and ACT prefer complex gestures to

an even greater extent than undirected events do (only 45.38% of directed incremental predicates were accompanied by simple gestures).

Except for model 3, all models predicted ended and simple gestures more accurately than continuous and complex gesture forms (Table 19).

	conditional inference trees			**random forests**			
	accuracy	correctly predicted		accuracy	correctly predicted		C
		e/simple	c/complex		e/simple	c/complex	
model 1	69.82%	76.32%	55.00%	73.78%	95.17%	25.00%	0.77
model 2	67.99%	88.89%	27.68%	71.95%	92.13%	33.04%	0.74
model 3	67.98%	65.94%	71.11%	74.12%	85.51%	56.67%	0.80
model 4	62.72%	81.62%	34.78%	67.11%	83.09%	43.48%	0.77

Table 19: *Model comparison*

5.4.3 Qualitative (micro)analysis

Participants of interactions occurring in a collaborative setting were shown to adopt various strategies regarding multimodal meaning-making. In a qualitative analysis based on the same material as the present study, Lehečková and Jehlička (2019) demonstrated that both English and Czech speakers, when talking about abstract entities, established a temporary shared gesture space between them (a process of *alignment*) where they maintained a shared gestural representation of the conceptual entity or entities, potentially adjusting it in a process of *elaboration*. The notion of elaboration was introduced by Langacker, who described it as a general cognitive process that may be involved at any level of linguistic representation and that involves a: "augmentation, adaptation, or further processing [of an already established baseline (B)] produc[ing] a structure that may itself function as B at another stage or level of organization" (Langacker, 2017, p. 239).

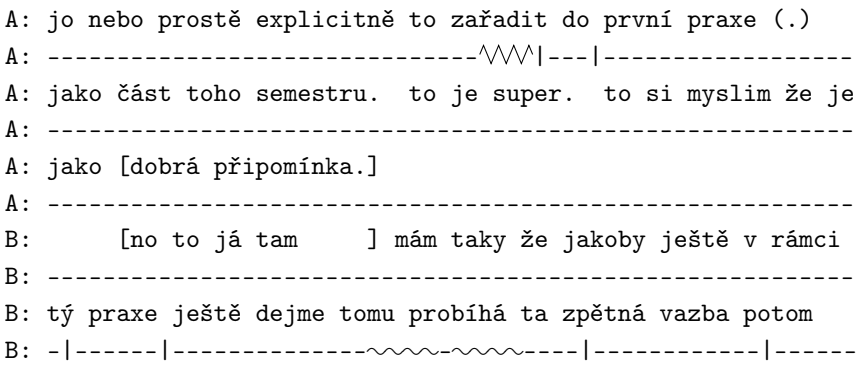
Turning to the interactional nature of multimodal construals of eventuality, a question that arises is whether and how speakers rely on intersubjective or, for that matter, intercorporeal cues signalled by the use of gestures, that may be associated with the two types of profiling introduced above. Let me briefly focus on two examples where the participants of the interactions share the attention to a conceptual entity and use gesture that are associated to its eventuality qualities. The examples are given as CA transcripts, supplemented with an additional line representing gesture – the notation of gestures is derived from the ELAN annotation (see above): [|] = *e*; [^^^|] = *ee*; [–~~~~–] = *cc*.

The first example (14), from the Czech subcorpus, captures two turns, in which the interlocutors A and B discuss practical training as part of a study programme. Speaker A suggests including a certain element into the training in order to provide

[96] There was no undirected – incremental combination in the Czech sample.

the students with feedback. Speaker B adds that some kind of feedback is provided after the training, but not during the training itself.

(14) CS021220, time 06:39 – 06:50



(translation: A: *yeah, or [we can] just* ***put it explicitly*** *to the first training, as a part of the semester, that's great, that's a good point*. B: *well, I also pointed out that during the training, they're, say,* ***collecting*** *feedback, afterwards*.)

The speaker A produces an ended complex gesture with ended subphases together with the verb zařadit (directed ACH). The speaker B produces a continuous complex gesture with continuous subphases together with the verb *probíhá* (undirected non-incremental ACT). In the latter case, the form of the gestures appears to be congruent with the construed aspectual contour of the event. In the former case, however, the multiplicity of the ended subphases does not seem to be in accord with the event structure. Let us take a look back, again, at example (5) from Chapter 2 (revisited in Chapter 4) from the same session, where the speaker A produced a gestalt-iconic gesture highlighting *dividing up* a training into phases. The gesture affiliated with the verb *zařadit* ('put into') refers to the very same conceptual object (the training) that is maintained in the shared gesture by recurrent gestural forms, profiling the temporal extent of the training as a timeline on horizontal axis. The complex gesture in question, in a form of repeated strokes, refers to the conceptual object several times. Rather than suggesting an iterative construal of the event signalled by repeated strokes, I argue that the repetition of the ended gesture is related to the contextual factors that come into play. First, the verb is modified by a "bounding" modifier *explicitně* ('explicitly'). Second, the immediately following complement *do tý první praxe* ('into the first training') is determined as well as involves a demonstrative pronoun. Even though these factors did not prove to be significant in the quantitative analysis, in this case, we should consider the occurrence of multiple factors at once. The association between determination or deixis and the *e*-gesture is likely the case with the complement phrase as contains a prosodic peak that might be linked to the last of the iterated gesture strokes.

The speaker B, using a complex unbounded gesture to highlight the event contour in line with the undirected non-increment ACT construal, also refers to the train-

ing timeline. With a bi-manual gesture, she first delimits a section of it (saying *v rámci tý praxe* ("as part of the training") and then highlights the contour of an event that takes place in the profiled section (*probíhá ta zpětná vazba*, "feedback is being collected"). This may be considered a case of elaboration of the BASELINE, i.e. the perpetually elaborated conceptual object *training*.

In the other example (15, the English subcorpus), two interlocutors talk about transcribing speech for a multimodal corpus, trying to figure out the interpretation of a particular segment of the transcription.

(15) EN2004a, time 23:01 – 23:12

```
A: so I think she just she didn't catch [the full form      ]
A: ---------------------------------------------------------
B:                                       [she was cutting off]
B: -------------------------------------------------------|--
B: rather than saying she would like (..)  to see something
B: ----------------------------------------~~~~-----------
A: right right yeah she I think meant to use the infinitive
A: ----------------------------------~~~~-~~~~|----------
A: but she just left off
A: ----------------|-----
```

The speaker B first produces a straightforward iconic "cutting" gesture while saying *cutting off* which could, perhaps, be seen as more of an embodied enactment rather than gestural marking of an event boundary. However, things become interesting when the speaker produces a *cc* gesture together with *to see something*. Such a sequence constitutes a clear contrast between a punctual (directed ACH/ACT) contour that is prototypical for the verb *cut* (and this prototypical contour is being profiled instead of focusing on the progress of *cutting* as an activity) and undirected ACT construal of *to see something*.

In the following turn by speaker A, a similar *cc*-gesture occurs with *to use the infinitive* (directed non-incremental ACT) and after that, we see an ended simplex gesture accompanying the verb *left off* (directed ACH). It cannot be ruled out that the use of the *cc*-gesture is to a certain degree primed by occurrence of a similar gesture form in the preceding turn. Such a view is supported by an observation, that the entire segment exhibits what can be called *gesture resonance*. Dialogical resonance (Haddington, 2007) was defined as "the ways in which interlocutors linguistically rely on previous utterances when producing their own talk" (Nir et al., 2014, p. 545), and is characterised by, among other means, repetition. Warner Garcia (2013) shown that resonance phenomena apply to the use of gestures too, not only in the form of simple repetition but also, and crucially, as a subject to collaborative elaboration (without a reference to Langacker's notion). A preliminary schematization of resonance manifested at the level of syntactic constructions as well as the gestural pattern, constituting a symmetrical multimodal resonance construction, is provided in Figure 35.

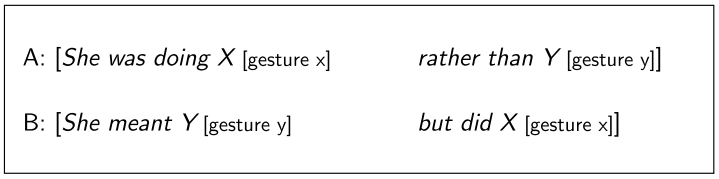


Figure 35: *The resonance construction*

5.5 Discussion

The quantitative analysis showed that the core evidence related to multimodal coding of aspectuality, namely that bounded (*ended*) and unbounded (*continuous*) gestures tend to pattern non-randomly with certain linguistic structures, attested in previous studies (predominantly based on experimental data and narrative tasks) is observable also in interactional data. However, the overall picture is a lot more complex due to the following factors. First and foremost, the dominant type of gesture across all delimited categories was an *ended simplex* gesture. Presumably, this prevalence of *e*-gestures results also from other factors than a particular aspectual constructional type. Among these, the fact that predicates and verbal phrases that were put under scrutiny in this study are often parts of a focus domain within the utterance (Lambrecht, 1994), and thus the simple ended gesture may serve (also) to profile the whole focus part of utterance information structure, and in order to fulfil this function, it may align with prosodic contour and sentence stress in particular (see Section 2.2), may be the most prominent explanation of the overall predominant representation of this kind of gestures regardless of the aspectual type. Results from other studies (see e.g. Cienki and Iriskhanova, 2018, for Russian) support the view that this type of distribution is not random, but due to other factors than mere aspectual structure. Given this overall distribution, it might, in fact, be the frequency of unbounded gestures that is indicative of aspectually motivated multimodal patterns. As English and Czech may, due to differences in their general aspectual systems, develop divergent multimodal patterns and therefore show significant cross-linguistic differences, let me first address the findings for the two languages separately.

As regards the gesture *outer boundedness* in English, it is the aspectual types and aspect that predict the distribution of ended versus continuous gestures – with ACH with the highest prevalence of *e*-gestures and ACT (realized by PRG forms) mostly attracting *c*-gestures. The multimodal pattern associated with ACH may be attributed to their internal structure: ACH take place in a very limited time span, which is either compressed to a mere point in time, as in cyclic ACH (*hiccup* (*once*)), or partially extended (in directed ACH, e.g. *Something happened there*) but relatively shorter than in

other aspectual types, especially ACC that are durative and directed. Thus, the relatively limited period during which ACH develop, seems to be the strongest trigger for the presence of ended gestures in English. The second finding, i.e. the patterning of *c*-gestures with ACT in progressive forms, is very much in line with the previous findings (Duncan, 2002; Parrill et al., 2013). Both of these results confirm the initial predictions. According to the model 1, it is also the complement number that explains the variation of outer boundedness types within non-ACH + non-PRG cluster (which supports the theoretical account of aspectuality encoding in English, cf. e.g. Filip, 1999). However, this might very well be a side effect of correlation between gesture boundedness and complexity: ended gestures tend to occur in simple forms whereas continuous gestures are typically complex – in this cluster, cases with plural complements had indeed higher proportion of complex gestures (0.62) compared to singular complements (0.39). Throughout the data, a pattern is apparent that gestural boundedness and complexity closely interact and it is often hard to tease the individual effects apart.

In the English dataset, gestural *complexity* is linked, on the one hand, to progressiveness and, on the other, to incrementality, which, in both cases, can be straightforwardly attributed to specific aspects of the corresponding aspectual contours. For activities expressed in progressive forms, it is the relative internal unboundedness of this construal that prompts the gestural profiling as the most prominent feature, the complex gesture being the proper choice for supporting this aspect. In this respect, Janda's (2003) metaphorical model of imperfective semantics as a fluid (see 4.3.1) entity aptly fits this multimodal pattern for English as well. Furthermore, the internal structure of incremental events consists in gradual development of the target event: gradual changes in the involved object (that is being created, destroyed or manipulated in other ways, e.g. *write a novel*, *eat lunch*, *paint a wall*) mapped onto gradual changes in an event that continue until the inherent endpoint of the event has been reached (or it remains implied in directed incremental ACT). Thus, the incremental development triggers the complex form of the accompanying gesture. For ACT construed linguistically via progressive forms, it is the relative temporal unboundedness of this construal that prompts the accompanying gestures signalling that in its nature, the target process could go on forever, that what matters (and therefore is profiled), is the flow of the process, and as such (profiled from within), it appears unlimited.

Anyhow, we must treat the results cautiously, as the models are not perfect representations of the data – their performance is relatively poor when it comes to predicting the occurrence of *c*-gestures and complex gesture forms (the difference in prediction accuracy between the random forest and the conditional inference tree models displayed in the Table 19 is mostly caused by the better prediction of ended and single gestures by the random forest models, while the prediction of continuous and complex gesture remained unimproved).

A different picture emerges from the Czech data. The grammatical distinction of PFV versus IPFV aspect correlates with the distribution of gesture boundedness types.

From the perspective of aspectual types, this also corresponds with the distinction between ACH and ACC on the one hand and ACT and STATE on the other, i.e. there are only marginal instances of *q*-unbounded (*atelic*) perfectives as well as only one type of tense-aspect-mood (TAM) constructional pattern of q-bounded imperfectives in Czech. Thus, on the one hand, This patterning seems to be in accord with previous claims in the literature (Filip, 1999, *inter alia*), namely that the aspectual dichotomy in Czech highlights the most fundamental difference in the internal structure of Czech events, prototypically correlating with telicity distinction, and thus will presumably be profiled by other expressive means, e.g. gestures. However, if we turn to the random forest model, we find the strongest predictors of gesture boundedness in the Czech sample within the lower-level aspectual features – the presence of the incrementality features increases the likelihood of *c*-gestures and the presence of the directedness feature significantly increases the likelihood of *e*-gestures. In directed events, it is the inherent endpoint that may prompt the ended gesture (and this holds also for directed ACT that bear an IPFV form). If a directed event is also incremental, the specific incremental internal structure may lead to gestural profiling of graduality through continuous (and presumably also complex, see below) gesture, regardless of the aspectual form. These findings are not contradictory, they simply turn a simplistic story into more complex one, supporting what Janda (2015) or Divjak (2011) have advocated: a single predictor can hardly be found of aspectual distribution, even in languages with grammatical aspect (e.g. Slavic). Those models have so far proved to be the best ones that consider (i) both conceptual frame and linguistic construal, and (ii) more categories beyond aspect, especially focusing on TAM patterns.

As for the gesture complexity, we should again be reminded that it correlates with gestural boundedness (in Czech data, 75.56% of *c*-gestures were complex). Bearing in mind the lower predictive power of the model, it indeed suggests the same pattern – with a prominent role of directedness and incrementality. Another possibly noteworthy pattern (although with limited evidence at the moment) emerges from the combinations of continuous and simple gestures. In Czech, there is a tendency of these gestures to accompany STATE predicates with a simple internal contour, referring to mental states, attitudes or intentions (such as *vědět* ('know'), *potřebovat* ('need') or *těšit se* ('look forward')). However, this pattern is based on only a limited number of instances and needs to be further attested on a larger sample.

In 19 cases, the annotators could not reliably attribute an aspectual type to a predicate without consulting the video material and the accompanying gestures; these instances were discarded from the final sample. From the perspective of multimodal

[96]Together with states, directed incremental ACT belong to a cluster with lowest *e*-gesture proportion – but it must be noted that they represent a marginal group (n = 5) in this cluster, yet coherent both formally and functionally (IPFV forms in the past tense denoting *q*-bounded events that incrementally developed to a natural endpoint which, contrary to PFV forms, is profiled by the past tense form, see e.g. *Ten dopis jsem posílal*.IPFV *už před dvěma dny* ('I've sent the letter two days ago')).

construal, such cases represent the above introduced *profiling-for-discrimination*, when gestures provide the critical cue required for the intended interpretation of an event. There were 3 types of ambiguities:

(i) ACT/ACH: *didaktici se musí trošku aktivovat* ('didacticians have to mobilize a bit');

(ii) ACT/ACC: *you can go and search the whole corpus*;

(iii) ACH/ACC: *musel absolvovat ty předměty* ('he had to pass the modules').

In English, the ambiguous cases included the types (i) and (ii) – in some contexts, the predicates allowed for both the ACT and ACH or ACC interpretation, i.e. the profiling of reaching an endpoint (telicity) depended on multimodal input. In Czech, the ambiguous cases were related to so-called biaspectual verbs (mostly lexical borrowings that do not formally discriminate between aspects (Chromý, 2014) – e.g. *absolvovat* ('to pass'). The type (iii) occurred only in the Czech sample.

While the 19 cases of represented only 3% of the data, these were only the clear instances of *profiling-for-discrimination*. Presumably, in the online language processing, the cases where are relatively frequent where gestures – as visually salient signals – provide the critical input for the interpretation of aspectual contour.

The qualitative analysis highlighted the interactional aspects of multimodal event construal, particularly the fact, that speakers build upon the shared common ground and make use of it in profiling event contours via collaborate meaning-making tools such as *elaboration*. Besides that, we saw that event representations may be primed or motivated by multimodal cues in the immediate context in the individual production as well as in dialogue. Adopting the interactional perspective, multimodal event construals can be approached with respect to a vast range of discourse phenomena such as gestural resonance. The qualitative microanalysis made it evident that no quantitative analysis can account for the multiplicity of factors that may underlie the choice of a particular gestural form in spontaneous production: given its limited scope, the primary aim of the qualitative analysis was to demonstrate what might be hidden behind the residual variance of any statistical model.

6. Experimental study

"Every decoding is another encoding" (Morris Zapp)

This chapter presents the second stage of the study – design and results of the comprehension experiment designed to validate the results of the corpus study introduced in the previous chapter. The experimental study builds upon the findings in Czech, namely on the observed association between grammatical aspect and the aspectuality feature of directedness and gesture form (profiling of event BOUNDEDNESS).

First, a brief overview of the experimental research of gesture or gesture-speech perception is provided, leading to identification of five issues related to collection of gesture perception data via behavioural experiments (i.e the subjects' responses to visually presented stimuli) and their analysis and interpretation. The rest of the chapter reports the design, methods and results of the experimental study. The chapter is concluded by an interim discussion.

6.1 Gesture perception: areas and methods of experimental research

Compared to the empirical study of gesture production, mostly based on elicitation of co-speech gestures accompanying spoken narratives, in the research area of how gesture functions in language comprehension processes a smaller number of empirical studies has been conducted. This imbalance is likely to be related to the limitations inherent to the behavioural methods of psycholinguistics, when it comes to co-speech gesture. A method widely applied in the gesture perception experiments is the "playback paradigm" where "the participant takes on the role of an observer, decoding the information they are presented with in the form of a video stimulus" (Holler, 2013, p. 839). The general methodological problem is how a phenomenon so complex and yet still only partially understood as co-speech gesture can be normalized to become an experimental stimulus to which a set of controllable variables may be attributed. This issue has been dealt with in various ways: one possible source of stimulus material are samples of natural gesture production (e.g. multimodal corpora or elicitation recordings made previously for the gesture production studies purposes or specifically for the comprehension studies). Elicited gesture production, however, is in general characterised by an unpredictable degree of variation in many respects and thus has only limited usability in comprehension research. The natural variation may be reduced when the gesture production samples are obtained through enactment, usually performed by researchers themselves or professional actors. This, in turn, comes with its

own price: gesturing may become noticeably unnatural, reducing the validity of the stimuli in the study of how gestures are perceived and comprehended in the natural communication. The same drawback applies to the cases when animated figures are employed to represent human gesture behaviour.

Nonetheless, a number experiments directed at gesture comprehension or perception that have been conducted so far provide a solid body of evidence. I will review the major lines of research in this area, focusing particularly on methodology.

Most of the studies that are of interest here are concerned with the integration of gesture and speech in the process of comprehension. This is not, however, always the case. For instance, one line of research addressed the priming effects of iconic gestures on lexical retrieval – contributing to the major agendas of psycholinguistics: a study of semantic priming[97] and the effects of iconicity in language processing. Bernardis et al. (2008) investigated the semantic relation between iconic gestures and words in a variant of the lexical recognition experiment and an ERP[98] study. In the lexical recognition, subjects (Italian native speakers) were presented with pantomimic gestures followed by written words either corresponding or not corresponding semantically to the silent gestures and were asked to read the word. The results revealed no priming effect for congruent condition, but when the gestures and words were not related, subjects' reaction times were significantly slower. The authors interpret this finding as being in favour of the Information Packaging Hypothesis (see Section 3.2.2). The subsequent ERP experiment supported the interpretation that the semantic encoding of gestures and words relies on different cognitive processes and revealed mutual interaction between gestural and lexical meanings, as manifested by the negative priming effect in the non-congruent condition.

Gesture priming was further explored by So et al. (2013) who ran a lexical decision experiment with co-speech gestures as primes (i.e. iconic gestures accompanied by matching words) or speech or gestures alone. Subjects' (speakers of Singapore English) reaction time were facilitated in all condition when the targets were semantically related to the primes. The most robust priming effect was observed in the co-speech gesture condition.

The divergent evidence of gesture priming effects provided by the two studies may result from the different tasks employed (lexical recognition vs. lexical decision), however such difference would be rather surprising. A more likely explanation may be related to the general methodological limitations of gesture perception experiments mentioned above. Although both studies used "iconic gestures" as primes, the way corresponding words were sampled is not consistent across the two studies, covering different semantic fields and levels of iconicity. It is thus too early to draw any conclu-

[97] Priming studies are based on a presentation of a *prime* stimulus and its effect on the processing of the *target* stimulus – positive or negative in terms of reaction time.

[98] ERP = event-related potentials, a method of measuring of the cortical neural activity based on the electroencephalography.

sions in this regard, as more evidence needs to be collected, let alone the need for the replication of the evidence that has already been put forward.

Studies focusing on the gestures co-occurring with speech frequently employ manipulation of the natural language production of some sort. The most basic kind of manipulation is separate presentation of speech with and without gesture. In one of the early empirical accounts of comprehension of gesture-speech integration, Beattie and Shovelton (1999) used questionnaires to ask their test subjects about semantic features of objects and actions addressed in narrations they had perceived in the form of audio-only or videos showing narrators gesturing. Subjects' answers were significantly more precise when gesture (in fact, visual modality as such) was present in the input. Their findings were further confirmed under more ecologically valid conditions: Holler et al. (2009) added other experimental conditions to the original design, including a setting in which subjects were exposed to the narration not in the videos, but when produced by a speaker physically present in (i.e. face-to-face communication). In the face-to-face condition, the precision of subjects' answer increased compared to co-speech gestures presented in videos. The authors conclude that this effect may be attributed both to social and pragmatic aspects of interaction with communicative partner and limitations of video-presentation (lower image quality and smaller scale of the gestures).

Interaction of meaning conveyed by gesture and speech was also investigated through mismatching the visual and acoustic information, either by stimuli containing gesture deliberately produced as incongruent with the content of speech, or by manipulation of video and audio signal. The former method was first introduced in a small-scale experiment by McNeill et al. (1994), who presented the participants with videos containing narratives pretended to be naturalistic but that were actually staged and involved mismatching gestures (as well as normal gestures). The mismatching gestures were manipulated to be incongruent with speech in terms of movement direction and temporal alignment. Compared to the subject group that was presented with audio-only version of the narratives, the participants subjected to the mismatching condition understood the story differently (as revealed in the subsequent re-tellings). This general design was later used by Özyürek et al. in their ERP study 2007 focusing on how gestural information is integrated with the meaning of spoken expression at syntactic and lexical level in Dutch. Videos of a speaker (actor) producing sentences with pre-defined gestures (congruent) were used as stimuli. The mismatch was achieved by extracting sound from the videos and a re-assembly of non-corresponding audio and video signals. A set of four experimental conditions (sentences with in/congruent gestures (to test lexical integration) as well as with semantically conflicting or congruent verbs (to test syntactic integration – see Figure 36)[99] was presented to the participants (16 speakers of Dutch).

[99] Source of the image: Özyürek et al., 2007, p. 608.

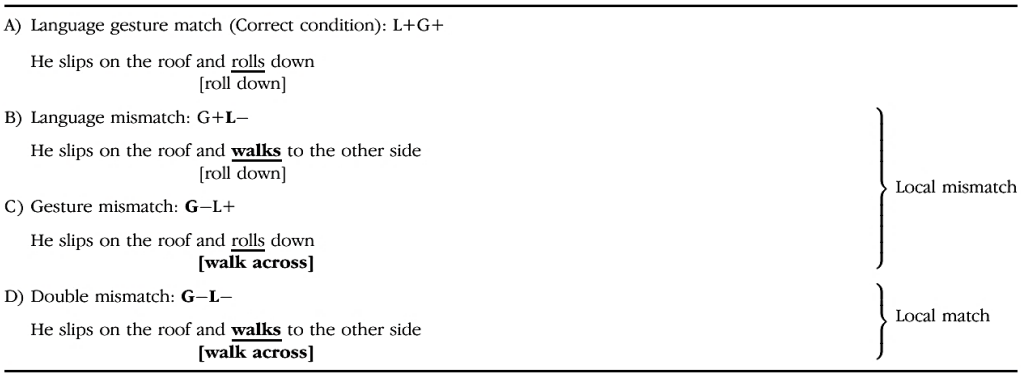


Figure 36: *Four conditions used in the experiment by Özyürek et al.*

All critical verbs were action or motion verbs and thus the corresponding iconic gestures were relatively simple to pre-define. The ERP measurement was not accompanied by any task, the subjects only watched the video stimuli. The results indicated that semantic information conveyed by gesture and speech is not only processed simultaneously (arguably employing the same the neural structures) in the case of their immediate co-occurrence in co-speech gestures, but also when processing mismatches on the syntactic level. These findings were supported by a fMRI study based on the same design (Willems et al., 2007).

Kelly et al. (2010) ran two experiments based on the mismatch paradigm. In the first experiment, 21 participants were asked to decide whether the presented gesture + action verb combination corresponded to a prime – a silent video clip with a person performing a simple action. Half of the target stimuli were mismatching – either with regard to speech or the gesture.[100] The error rates were found to be significantly higher in the mismatching condition (more so when gesture was incongruent). In the subsequent experiment, a different group of 42 subject focused only on the speech. The results revealed that the more incongruency there was between gesture and speech in the targets, the higher (significantly) was the error rate. According to the authors, this demonstrates that "gesture and speech interact in a mutual and obligatory fashion, and when conveying the same message, they greatly enhance understanding" (Kelly et al., 2010, p. 266). Based upon this general observation, they formulated the Iintegrated Systems Hypothesis assuming bidirectional interaction between gesture and speech during spoken language processing (see Section 3.2.2).

Another line of research tackled the role of gesture in speech disambiguation or discrimination. In their ERP study, Holle and Gunter (2007) presented the subjects (27 German native speakers) with complex sentences beginning ambiguously due to a homonymous words in the initial clause (e. g. *Sie kontrollierte den Ball* → *was sich*

[100] For example, if the prime showed a person cutting vegetables, the gesture-mismatching target would contain the word *chop* accompanied by a twisting gesture, the speech-mismatching target would contain the "twist" accompanied by the chopping gesture.

im Spiel beim Aufschlag deutlich zeigte / Sie kontrollierte den Ball → was sich Tanz mit dem Bräutigam deutlich zeigte)[101] which were accompanied by gestures (performed by an actor) iconically depicting the profiled meaning. In half of the stimuli, video and audio were mismatched so that gesture profiled the incorrect meaning of the ambiguous word. Participants judged the congruence of the gesture with the target meaning. Behavioural as well as ERP results suggested that speakers indeed take advantage of iconic gestures when decoding ambiguous speech input, especially so when the target meaning is the less frequent of the two competing meanings.

In their ERP study, Drijvers et al. (2018) explored the role of gestures in decoding information when the speech signal is obscured by noise. In their experiment, 29 Dutch natives were presented with videos showing a speaker producing isolated words – verbs of action – either together with a co-speech gesture or without. The speaker recorded for the stimulus material was again an actor, instructed to carry out iconic gestures along with half of the verbs. Half of the stimuli contained manipulated sound channel (voice distorted using vocoder), leading to total of four stimulus groups: speech-only with clear sound or degraded × with iconic gesture or without. Participants' task was to identify the last verb from the series of previously seen stimuli from the list of four options including phonologically and semantically related and unrelated verbs (*cued-recall task*). The assumption that gesture will increase the success rate in the task in the degraded speech condition was confirmed. When decoding the degrading speech, participants were significantly more likely to recall the correct verb when the verb was present. Besides that, the reaction times were significantly faster in both sound conditions when gesture was present. The ERP results pointed to right superior temporal sulcus, suggesting its role in exploiting the visual cues when perception of acoustic information is difficult: "listeners might engage their motor cortex to possibly simulate gestures more when speech is degraded to extract semantic information from the gesture to aid degraded speech comprehension" (Drijvers et al., 2018, 11).

The comprehension of the expression of eventuality in speech and gesture was addressed experimentally by Becker et al. (2011). Following up their production study (see Section 4.2.2), the authors conducted a comprehension experiment to test whether the patterns they discovered in the recorded narratives (i.e. ACH events being marked with bounded gestures) play a role in comprehension too. Fourteen extracts from the recordings collected in the production study were presented as stimuli to 26 native and L2 speakers of English. Half of the stimuli consisted of videos depicting bounded gestures accompanying an ACH expression, the other half of the stimuli was manipulated – videos with gestures corresponding to ACT or ACC were assembled together with sound tracks containing ACH expressions. To ensure that the mismatch between video and audio would not be apparent, the videos were edited not to display speakers'

[101] 'She controlled the ball → as become evident in the game/dance'.

heads. The participants' task was twofold. First, they decided whether a verb displayed on the screen after the stimulus had been presented is the verb they heard, while reaction time was measured. Then, the participants were asked to judge whether the video was edited or not and their answers were used to assess the degree of subjective sensitivity to manipulation with the stimuli. The results revealed the both low- and high-sensitivity subjects' success rate was significantly lower in the mismatch condition. The effect of gesture-speech incongruity was also present in reaction times of the subjects with a relatively low sensitivity to stimulus manipulation – they were significantly slower when gesture and speech did not correspond. The comprehension experiment thus supported the findings reported in the production study (results of both parts of study are discussed in Section 4.3). In the multimodal expression of event structure – linguistically encoded in the lexical semantics of the verb and visually encoded in the form of an iconic gesture – certain features (in this case, telicity) tend to be reflected by co-speech gestures' form (in this case, boundedness) in spoken production. These multimodal eventuality constructions are perceived as units – whenever the expected linkage between the two modalities is disrupted in communication, the addressee's comprehension might be affected.

One of the interesting methodological aspects addressed by this study is measuring of the sensitivity to stimulus manipulation. The authors offer alternative interpretations of the effect of sensitivity (quantified as *d'* statistic (Swets et al., 1961) discussed in more detail below) to editing, as it is not clear what exactly could the sensitivity level be manifestation of. As the low-sensitivity subjects had longer reactions in mismatching condition than high-sensitivity subjects but at the same time they had faster reactions in matching condition, it is not the case that sensitivity to manipulation also means the sensitivity to event structure mismatch. On the contrary, the subjects with high reaction times in mismatch conditions and low sensitivity scores might in fact be "overinterpreting" the mismatch condition as plausible, which costs them the additional reaction time. This illustrates the fact that controlling the effects of stimuli manipulation and correlating them with the actual comprehension of the communication content is not at all straightforward. However, this may be dealt with via post-hoc questioning the subjects about their strategies (rather than by implicit testing).

This review is of course not comprehensive. Holler (2013) provides a concise overview of the field of gesture comprehension research, also with respect to the earlier studies. And again, I leave aside completely two major research areas here – the research of gesture in the context of L1 acquisition and the research of gesture in speakers with language disorders. Apart from that, I also cannot pay appropriate attention here to the neurolinguistic studies of co-speech gesture. This field was thoroughly reviewed by Marstaller and Burianová (2014).

Experimental studies of gesture comprehension have addressed numerous topics, however the evidence yielded does not allow us to make a synthesis that would

support any of the existing psycholinguistic models of language comprehension. What we can deduct from it, though, aligns with the general findings of the research of the role of iconicity in language comprehension. In short: every iconic cue available cue is exploited in comprehension.

From the methodological perspective, the research of gesture perception has faced many challenges and many of them remain unresolved. As noted at the beginning of this chapter, most of the challenges are related to the general paradigm of behavioural research. If speech itself is not easily reduced to experimental stimuli, in combination with visual modality it becomes an incredibly information-rich signal (Perniss and Vigliocco, 2014). The number of possibly involved factors that need to be accounted for (as explanatory variables) is extensive while only a fraction of them can actually be controlled. In the above review, we have seen some of the recurrent issues. Listed below are these that I consider crucial:

(i) *ecological validity of gestural stimuli*

For reasons described above, enacted gesture production is frequently used for comprehension experiments. As it is arguably more difficult – even for a professional actor – to simulate spontaneously produced gestures than to modulate speech for the purposes of a given research, one has to be particularly wary of to what extent is it legitimate to consider the gesture sample a representation of a "natural" gesture. This may be adequately dealt with via conducting acceptability ratings with a separate group of participants before using the stimuli with the experimental subjects.

(ii) *degree of inclusion of other types of bodily behaviour*

The reviewed studies vary regarding inclusion of other types of non-verbal expression, such as head movement, facial expression or eye gaze. Yet, any kind of bodily behaviour may become a part of multimodally conveyed information and therefore has to be controlled for within the experimental design. Focus on a particular aspect of co-speech gesture phonology, e. g. handshape or movement dynamics may be induced in subjects either by obscuring face or head of the speaker visible in a stimulus video, or by editing the non-focal parts of the speaker out of the video. The latter solution may be preferable considering the ecological validity of the stimulus material.

(iii) *scalar vs. discrete view of gestures*

Co-speech gestures in general tend to be continuous phenomena rather than discrete units. While this is certainly true considering the linear segmentation of co-speech gestures along the stream of spoken production for the studies reviewed here, the more important thing is that co-speech gestures cannot be easily labelled as belonging to a single semiotic or functional category, e.g. "iconic

gesture". Such a broad (and constructed) category should therefore be avoided when delimiting the stimuli. Rather, more specific semantic categories combined with specific functions (such as MANIPULATION, MOTION-MANNER, or SHAPE.POINTING (gesture iconically depicting shape and at the same time used deictically) and optionally also with phonological features should be preferred in description of the stimuli as well as in the instructions for actors/gesturers.

(iv) *scalar vs. discrete view of iconicity*

When referring to iconicity, one must be aware that the way it is manifested in gestures, speech or sign may have many various forms. The instances of iconicity vary in the degree of similarity between the form of the expression and the physical appearance of the referent on the one hand and in the mode of representation on the other. The notion of gradual iconicity has been adopted within sign linguistics and recently it found its way to gesture studies as well – Chapter 2 covers this issue in detail. The recent studies by Hassemer and Winter (2018) or Ortega and Özyürek (2019) provide instructive examples of how to deal with gestural iconicity from a scalar or multidimensional perspective.

Although it is argued here that these issues represent a bundle of challenges that gesture researchers must face before any study is conducted, it could very well be said that the problems discussed above apply to any linguistic study based on a behavioural paradigm and using samples of natural language production to study comprehension.

6.2 The present study

Building upon previous studies (primarily on Becker et al., 2011, and Becker and Gonzalez-Marquez, 2018) while trying not to fall into the methodological traps discussed above, the experimental study presented here is assisted by a combination of data sources that have not been exploited (in this particular combination) before in gesture-speech comprehension experiments: the stimulus material used in the present study was designed with the help of (i) *corpus data*, (ii) native-speaker linguistic intuition (data from a special-purpose *rating study*) and (iii) *motion capture* (MoCap) data.

The experimental part of the study focuses only on the Czech speakers' comprehension of multimodal constructions. In particular, it focuses on speaker's judgements concerning combinations of *ended* and *continuous* gestures with sentences with PFV or IPFV predicates.

The corpus study (Section 5.4, forest and tree model 3, Figure 33) showed that, in the Czech sample, *incrementality* and *directedness* are the most important predictors of gesture outer boundedness, but in general, it is also the aspectual dichotomy that is associated with the presence of an accentuated ending in gestures. As the majority of the *c*-gestures was also complex, it is assumed that the correlate of incrementality

	directedness		
outer boundedness	directed (n = 138)	undirected (n = 57)	NA (n = 33)
ended (138)	98 (0.71 / 0.71)	29 (0.51 / 0.21)	11
continuous (90)	40 (0.29 / 0.44)	28 (0.49 / 0.31)	22

	aspect	
	PFV (n = 104)	IPFV (n = 124)
ended	76 (0.73 / 0.55)	62 (0.50 / 0.45)
continuous	28 (0.27 / 0.31)	62 (0.50 / 0.69)

	aspect*directedness			
	PFV		IPFV	
	directed (99)	undirected (5)	directed (39)	undirected + NA (85)
ended	72 (0.73 / 0.52)	4 (0.80 / 0.03)	26 (0.67 / 0.19)	36 (0.42 / 0.26)
continuous	27 (0.27 / 0.30)	1 (0.20 / 0.01)	13 (0.33 / 0.14)	49 (0.58 / 0.54)

Table 20: *Associations between gesture outer boundedness and aspect and directedness in the Czech subcorpus*

(The numbers in parentheses represent proportions – left = column-wise, right = row-wise)

in gestural form would be complexity (a multiplicity of movement phases highlighting the incremental phases of the event) rather than the boundary marking. Table 20 presents the distribution of *e*- and *c*-gestures accompanying un-/directed and IPFV/PFV predicates in the Czech subcorpus.

Undirected PFV included cyclic ACH (*plácnout*, 'clap', *objevit se* 'suddenly appear') that are expected to be strongly associated with *e*-gestures – however, there were only five instances of undirected PFV in the Czech sample altogether. Undirected IPFV, on the other hand, were only represented by heterogeneous activities, *q*-unbounded and with an unspecified internal event structure. About a half of the *c*-gesture-accompanied IPFV was represented by STATE predicates to which the directedness property does not apply.

Looking at the interaction between aspect and directedness, 52% of *e*-gestures occurred with directed PFV and 54% of *c*-gestures co-occurred with undirected (or STATE) IPFV.[102] Directed IPFV were mostly accompanied by *e*-gestures (67%).

In sum, directed events were associated with bounded gestures regardless of aspect, undirected events were strongly associated with unbounded gestures only in IPFV.

The present experiment investigates to what extent Czech speakers perceive combinations of an *e*-gesture and a directed event and a *c*-gesture and an undirected event encoded in IPFV form as multimodal constructions. One of the ways to tackle this question is to explore the speakers' sensitivity to the violation of the assumed multimodal constructions, i.e. their judgements of "multimodal incongruencies", i.e.

[102] STATE accounted for 33 IPFV predicates, 52 were undirected, 39 directed. See Table 20.

directed predicate + *c*-gesture or undirected event + IPFV + *c*-gesture.

Three types of data will be used to analyse the speaker's sensitivity to multimodal constructions:

(i) participants responses in a forced choice task based on presentation of two sentences differing in aspect/directedness, otherwise identical;

(ii) reaction times required to perform the task;

(iii) participant response strategies reflected by the measures from the Signal Detection Theory framework.

As the corpus study clearly demonstrated, multimodal construal of eventuality is a multifarious process embedded in the context of the speaker's own discourse as well as the context of the interaction between other speakers and the physical surroundings. A reasonable way thus must have been found to transform the complexity of multimodal production into stimuli for a behavioural experiment. A close attention was paid to the preparation of both linguistic and gestural sides of the stimulus items, using a number of techniques to control as many variables related to the stimulus materials as possible. Maintaining the naturalness of the stimulus items while keeping the number of the degrees of freedom as low as possible was the primary objective during the preparation of the experimental design. On top of two pilot studies conducted prior to the final data collection, the design was subjected to several rounds of informal peer discussions during which the stimulus material was reviewed and elaborated.

6.2.1 Methods

As in Becker et al.'s (2011; 2018) comprehension experiment, it would be ideal for the direct validation of the corpus study to sample the excerpts of gesture production that would serve as stimuli in the behavioural experiment from the very corpus explored in the production part of this study. However, in the case of the Czech subcorpus described in Chapter 5, this path could not be taken for a number of major reasons. As the predicate is the linguistic unit which is in focus of this study, and not an isolated verb itself, it is virtually impossible to obtain a sample of instances from the corpus that would not be highly heterogenous in terms of, e.g. syntactic complexity, frequency, or semantics of its complements. Apart from that, the interactional setting of the corpus makes it often difficult to simply extract a predicate without losing intelligibility of the segment. Crucially, the enormous diversity in gesture form itself, combined with the varying camera angles, add up to an obstacle far too big.

Given the nature of the corpus at hand, I gathered recordings of gesture production by an instructed actor, bearing in mind the problematic issues related to this method mentioned above. The design of the stimuli and the procedure of the recording itself is described in detail below.

Another divergence from the previous experimental studies of gesture and eventuality considers the nature of the behavioural task and the way the experiment was distributed.

The presentation of stimuli did not involve sound – this was not in fact a methodological novelty but a choice to avoid the degrees of freedom introduced by speech, including controlling the prosodic realization or duration of each item.

Unlike the previous studies, the data for the present study were collected using an online testing environment. Online experiments introduce a whole new range of technical and methodological challenges, yet their advantage over the standard lab-based experiments is not insignificant – they allow for a more diverse sample of population beyond the over-represented population of college students.

Finally, a novel aspect of the experimental design is the use of a *forced choice* paradigm: rather than asking the participant to recognize the verbs and thus driving their focus on the lexical items, the task was based on a choice between two sentences according to their compatibility with the presented gestures. Participants of the experiment were presented with video clips of a person producing a gesture together with an utterance. The videos contained no sound and the face of the speaker was blurred. After each clip, two sentences were displayed on the screen, and the participants were asked to decide which sentence belonged to the video they had just seen. Reaction times and participants' responses were recorded.

Stimuli

The experimental study involved two kinds of stimulus items: gestures and sentences. In order to test the assumptions about multimodal constructions with PFV and IPFV verbs, the idea behind the design of the stimulus sentences was that they should represent "aspectual minimal pairs", i.e. pairs of identical sentences that differ only in the verb's aspect.

Stimulus sentences

Prior to construction of the stimulus sentences, a list of verbs was selected in a two-stage process. First, a list of one hundred most frequent transitive verbs (regardless of aspect) was obtained from the Czech National Corpus (Křen et al., 2016, version Syn2015)[103] a reference corpus of written Czech. After the exclusion of verbs that were not transitive, biaspectual verbs and *(im)perfectiva tanta*, the list consisted of 61 verbs.

A vast majority of the verbs (54) was attested in the PFV form, only 7 of the verbs in the list occurred more frequently in the IPFV form. The complete list can be found in Appendix E.

[103]The following CQL query was used: `[V.....R.*]`.

The second stage of the selection process considered what I call the *gestural affordance* of the verbs, i.e. the fact that meanings of certain verbs afford representation in visuo-motoric modality easier than others. In the context of this study, it is crucial to eliminate the interference of the verbs' semantics: for instance, the verb *to cut* in combination with an *e*-gesture could be interpreted as plausible regardless of aspect. To assess the gestural affordance of the selected verbs, a rating study was conducted with the native speakers of Czech, who were asked to rate the *imageability* of the verbs on a scale. A number of established subjective lexical norms lend themselves to be employed as a proxy for gestural affordance, including *iconicity* (Perry et al., 2015), *manipulability* (Masson-Carro et al., 2017) or *imageability* and *concreteness* (Paivio et al., 1968; Beattie and Shovelton, 2002). Imageability ("the extent to which [a word] evokes a mental image", Paivio et al., 1968, p. 79) akin also to sensory experience (see discussion in Chapter 2), defined as "the degree to which words evoke a sensory or perceptual experience" (Juhasz and Yap, 2013), represents a concept that is best applicable to verbs. For the purposes of the present study,[104] imageability and sensory experience were blended into a single norm (henceforth referred to a as imageability).[105]

The list of 61 verbs was distributed online via *Google Forms* (see Appendix D) to 12 raters – native speakers of Czech (6 women, 6 men, average age 33, age range [25; 47]) – who were provided with the following description: *Verbs vary in the degree to which they evoke certain sensory experience, i.e. an image associated with a process, activity or an event expressed by the verb, an image that does not necessarily have to be visual ("mental image"), but also auditory, olfactory or motoric.* The participants were asked to rate the verbs on a 5-point Likert scale (1 = "the verb does not evoke a sensory experience", 5 = "the verb instantly or very easily evokes a vivid sensory experience"). The verbs were presented in the aspectual form that was found to be more frequent in the corpus.

Table 21 presents the verbs with highest and lowest mean imageability rating. The five least imageable verbs belonged to *verba cogitandi* or verbs of mental states. The verbs with high imageability ratings were all verbs of physical action or object manipulation.

Twenty-four verbs that were rated as the least imageable were selected to the final sample and are listed in Table 22 in the PFV from and ordered by the relative frequency of PFV measured as the number of instances per million tokens (*ipm*).

The majority of the verbs (20 out of 24) in the sample occurs more frequently in the PFV form. The difference between the mean relative frequency of the PFV (225.66) and the IPFV (112.25) variants of the 24 verbs was significant ($V_{(23)} = 252$, $p = 0.001$ (Wilcoxon test, paired, one-tailed)).

[104] The available collection of lexical norms for Czech, including concreteness, specificity and imageability of 35 verbs (Kříž and Smolík, 2015) was not suitable for the purpose of the present study, because it covered only a fraction of the sampled verbs.

[105] Recall (Chapter 2) that Winter et al., 2017, also found sensory experience to be highly correlated to iconicity.

lemma	translation	mean imageability	SD
znát	'know'	1.08	0.29
zjistit	'find out'	1.17	0.39
nechat	'leave' or 'let'	1.25	0.45
rozhodnout	'decide'	1.25	0.45
zažít	'experience'	1.27	0.65
	…omitted…		
otevřít	'open'	4.17	0.72
vstoupit	'enter'	4.17	0.58
zavřít	'close'	4.17	0.94
psát	'write'	4.42	0.79
hodit	'throw'	4.50	0.90

Table 21: *Verbs with the lowest and highest mean imageability scores*
SD = standard deviation

lemma	translation	ipm (PFV)	ipm (IPFV)	PFV – IPFV
začít	'begin'	816.71	204.8	611.91
najít	'find'	507.31	133.46	373.85
udělat	'make'	497.02	625.10	-128.08
nechat	'leave'	453.73	54.24	399.49
získat	'gain'	307.90	41.42	266.48
rozhodnout	'decide'	276.63	54.64	221.99
zjistit	'find out'	220.62	25.66	194.96
změnit	'change'	217.61	134.32	83.29
připravit	'prepare'	193.08	78.38	114.7
vydat	'issue/give away'	175.74	54.87	120.87
dodat	'provide'	173.67	95.12	78.55
vytvořit	'create'	171.94	101.86	70.08
poznat	'get to know'	159.17	342.22	-183.05
zapomenout	'forget'	153.11	29.94	123.17
vybrat	'choose'	148.48	59.13	89.35
pochopit	'understand'	135.70	158.41	-22.71
vysvětlit	'explain'	131.74	117.90	13.84
ztratit	'lose'	129.49	60.98	68.51
vyhrát	'win'	120.70	14.01	106.69
nabídnout	'offer'	115.10	198.35	-83.25
opustit	'abandon'	98.16	25.42	72.74
odmítnout	'refuse'	79.68	60.62	19.06
poznamenat	'point out'	67.53	5.60	61.93
zažít	'experience'	64.94	17.53	47.41
mean ipm:		*225.66 (SD = 177.71)*	*112.25 (SD = 133.75)*	

Table 22: *Stimulus verbs*
English translation here is illustrative: the aspectual variants of the Czech verbs should be better translated to English as different lexical units

The exemptions are the aspectual pairs *dělat – udělat* ('to make' or 'to do'), *znát – poznat* ('to know' / 'to get to know'), *chápat – pochopit* ('to understand') and *nabízet – nabídnout* ('to offer').Except for the *nabízet – nabídnout* pair, these cases also differ from the rest of the verbs in the sample in terms of word-formation. The PFV forms *poznat*, *pochopit* and *udělat* are derived from the IPFV by prefixation, while the remainder of the verbs is charecterised by the aspectual variation in suffixes. The main issue with the prefixed PFV forms is that the prefix is not only a marker of perfectivity but can introduce additional semantic features of lexical aspect: prefix *po-*, for instance, may convey either distributive or delimitative meaning. According to the traditional view (e.g. Kopečný, 1962), some prefixes, including *u-*, do not carry any Aktionsart meaning and are considered "purely aspectual". A common test for such cases involves a possibility of forming a secondary IPFV by from the prefixed PFV form. When a secondary IPFV form cannot be formed, e.g. *udělat*.PFV > **udělávat*.IPFV or *pochopit*.PFV > **pochopovat*.IPFV,[106] the prefix is considered to be purely aspectual (but not across all lexical item, cf. *poznat*.PFV > *poznávat*.IPFV).

This view, however, has been disputed (see, e.g. Veselý, 2014). In her analysis of Russian prefixes based on grammatical profiling, Janda (2013b) suggested that all aspectual prefixes carry a specific meaning, but in some cases, there is an overlap between the semantics of the prefix and the base verb.

To take the possible semantic shifts in the prefixed PFV forms in account, a similar grammatical profiling should be carried out with the Czech verbs, which is beyond the scope of this study. Taking the cognitive perspective, I assume that the semantic relation within every aspectual pairing is a matter of construal, regardless of derivational process. Also, in various grammatical profiles or contexts, alternate aspectual counterparts may occur: *Ty údaje jsem hledal*.IPFV/*vyhledal*.PFV vs. *Ty klíče jsem hledal*.IPFV/*našel*.PFV (Starý Kořánová, 2019).

Given the complexity of relations in the aspectuality domain illustrated above, it is practically impossible to construct a sample of aspectual pairs that do not exhibit some degree of semantic variance and, at the same time, meet the lexical semantic (gestural affordance) and usage-based (frequency) requirements of the present study.

To ensure that the gestures would be consistently affiliated with the verbs, the sentences had a verb-final syntactic structure. All items were simple transitive sentences with SOV or OSV word order and in the past or present tense. The PFV sentences were all directed ACH or directed non-incremental ACC, the IPFV included directed as well as undirected ACT and STATE. The examples below illustrate the two types of sentence pairing: PFV_{dir} – IPFV_{undir} (16) and PFV_{dir} – IPFV_{dir} (17).

(16) A (IPFV): *Aritmetiku jsem dobře chápal.* —— B (PFV): *Aritmetiku jsem dobře pochopil.*

[106] The IPFV form *chápat* has another aspectual counterpart *chopit* ('to seize'). In modern Czech, however, the verb *chopit* is used only in the reflexive form *chopit se*, which is archaic and relatively infrequent (ipm = 10.98).

('I had a good grasp of arithmetic / I understood arithmetic')

(17) A (PFV): *Ten zápas jsme vyhráli.* —— B (IPFV): *Ten zápas jsme vyhrávali.*
('We won / were winning the match')

However, most of the verbs in the sample did not allow for an aspectual pair that would also differ in directedness: it is only the case with three sentence pairs.

In order to account for the possible effect of the predictors that were found important in the production study, the stimulus items varied in terms of the object number and deixis (here operationalized as a determination of the object by a demonstrative pronoun).

As distractors, 14 sentence pairs were used with the same syntactic structure as the critical items, but with two different verbs (partly selected from the high-imageability verbs from the rating study). The complete list of stimulus sentences is provided in Appendix F.

Videos

The sample of gestures used in the stimulus videos was produced by myself, following the operationalization of gestural formal features based on Bressem's guidelines 2013. A set of 63 videos capturing production of a single gesture was recorded, divided into five different formal groups (Table 23).

condition	**label**	**orientation**	**position**	**movement**	**handshape**
a	"cutting"	palm-lateral toward center	upper right-center	downward away-body accelerated accentuated ending	flat hand
b	"closing"	palm-lateral toward center	upper right-center	downward away-body accelerated accentuated ending	flat hand → fist
c	"cyclic"	palm-lateral toward center	upper right-center	downward away-body spiral	flat hand
d	"fading"	palm-lateral toward center	upper right-center	downward away-body accelerated	flat hand
e	"enactments"	variable	variable	variable	variable

Table 23: *Five conditions of gesture production (stimulus material): phonological operationalization*

Conditions (a) and (b) represented the ended gestures, conditions (c) and (d) represented the continuous gestures and condition (e) comprised the material intended to be used as the filler items (distractors).

Gestures in conditions (a)–(d) were produced in sets of 12 gestures varying only in the degree of acceleration and extension of the stroke phase.

Condition (e) comprised 15 ostensibly iconic gestures corresponding to 15 verbs – 6 high-imageability verbs from the rating study and 9 additionally selected verbs that were presumed to have a high affordance for iconic representation by enactment – all were action verbs, most of them were the verbs of human action, some involving handling a tool.

The critical gestures were performed by the right hand and so were the filler ones except for the two-hand gestures.

Recording of the video stimuli was carried out with the help of a motion-capture system in the Natural Media Lab at the RWTH University in Aachen. For the motion-capture, a Vicon Nexus (v. 2.9.2) system of infra-red cameras recording at frame rate of 100 fps was employed. The sensors (markers) were placed upon gesturer's right index finger (rIF), right wrist, right elbow, chest, left elbow and left wrist. The videos were recorded on a Sony digital camcorder in MP4 format using 1280×720 px resolution at 25 fps.[107] The gesturer was seated, and the camera was placed at about 30° off the subject's sagittal axis to his right-hand side (see Figure 37).

Figure 37: *Setting of the stimulus videos*
The white lines highlight the position of the MoCap markers (grey dots).

The video clips were edited in *Shotcut*[108] an open-source video editing software. The

[107] The video clips are available at the OSF repository.
[108] https://shotcut.org

gesturer's face was blurred and the audio track was removed. Blurring the face instead of cropping the video so as to display headless figures (cf. Becker et al., 2011) has one major advantage: traces of head and jaw movement remain visible in the video stimuli – obscured enough not to provide the receiver with clues for interpreting the speech, but apparent enough to make an impression that the gesturer was actually speaking.

The primary motivation for the employment of motion-capture was to provide an exact assessment of the phonological parameter in question: the slope of acceleration and deceleration. To ensure that the gesturer's rendition of an ended (simplex) gesture in contrast to a continuous (simplex) gesture really involved a clearly distinguishable difference in the execution of the stroke phase. Figure 38 displays the horizontal position (measured as distance from the baseline in mm) of the RH-IF marker during the execution of an ended gesture (blue line) and a continuous gesture (red line).

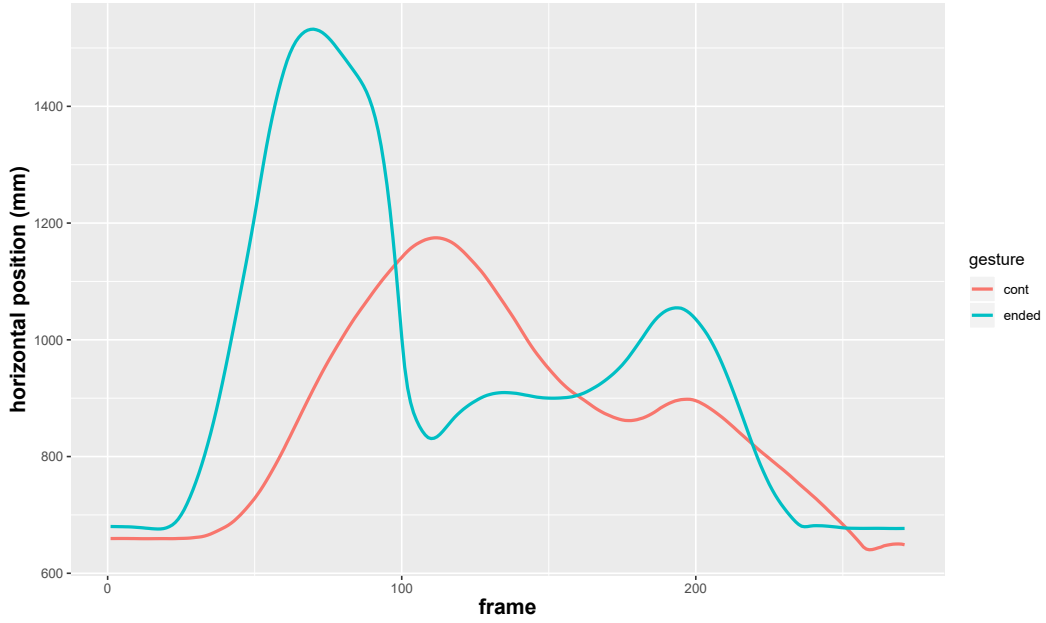


Figure 38: *Horizontal position of the RH-IF marker during the stroke phase*
Continuous (red line) vs. *ended* gesture (blue line) superimposed

The peak marks the onset of the gesture stroke phase, ending in the lowest point of the valley. In case of the ended gesture, we can see a very steep and rapid progress of the stroke phase, abruptly ended and followed by a brief post-stroke hold. The stroke phase of the continuous gesture, on the other hand, is characterized by a gentle downward slope ending in a relatively shallow valley – the transition to the retraction phase is smooth, without a visible interruption.

To eliminate the possible factor of phonological variation, one example of an *e*-gesture and one example of a *c*-gesture were selected as gestural stimuli in the critical conditions (the ones visualised in Figure 38 above, "cutting" and "fading").

Procedure

The experiment was created in *PsychoPy3* version 3.2.4 (Peirce, 2007), a Python-based stimulus presentation software, and was administered online through *Pavlovia*[109] a research platform hosted by the University of Nottingham, that allows for running and distributing online studies designed in PsychoPy.[110]

Participants completed the experiment on their own devices. Upon starting the experiment, the information about age, gender, native language and handedness was collected from the participants. After completing the questionnaire, they were provided with written instructions. They were informed that the videos will show a person "uttering a sentence" but will not contain sound and that their tasks will be to "decide which or the presented sentences belongs to the video". Before the experiment itself, the participants familiarised themselves with the user interface during three practice trials (see Appendix G for illustrations of the user interface and the text of the instructions).

In each experimental trial, a video was played, followed by a blank screen displayed for 500 ms, followed by a screen with two sentences displayed parallel to each other. The order of PFV and IPFV sentences was random. Reaction time clock was activated at the moment of the sentence display and the subjects answered using left or right arrow keys, according to the spatial order of the presented stimulus sentences. Subjects were instructed to answer as quickly as possible while the sentences were displayed. As soon as the response was recorded, the following trial was commenced automatically. There was no time limit for the response, but after 10 seconds, the sentences disappeared. Reaction times and the response key were recorded.

One session took approximately 10 minutes. Mid-session, a break window was displayed, informing the participants about the progress of the experiment (feedback on their performance was not provided). After completion of each session, the results were saved on the Pavlovia server.

Participants

Altogether, data from 40 participants were collected – 26 female, 13 male, 1 of an undisclosed gender, with a mean age of 28.43 years (ranging from 16 to 63 years). All subjects were native speakers of Czech. Link to the experiment was distributed via a network of contacts instructed to provide it to their families and friends who were native Czech speakers without a formal training in linguistics.

[109] https://pavlovia.org

[110] The entire documentation for the present experiment is available at https://gitlab.pavlovia.org/jakub.jehlicka/forcedchoice2.

6.2.2 Analysis and results

Two multiple regression analyses were carried out to investigate (i) the predictors of the aspectual choice (at this point understood formally as a choice of a stimulus sentence which involved an aspectual difference, without presupposing any kind of participants' strategy – that will be discussed below) and (ii) the factors influencing the participants' reaction times. As the factors in the model of aspectual choice, a selection of the variables from the corpus study (Chapter 5 was used: *outer boundedness* of the stimulus gestures, predicate *directedness* of the IPFV sentence, *deixis* and *object number*. The model for the RT data included also *aspect* as predictor as well as an additional variable: *PFV-IPFV frequency ratio*.

Aspectual choice

After the data reduction (see below), 937 responses were included in the quantitative analysis. Proportions of the response types[111] (IPFV vs. PFV sentences) in the two experimental conditions (continuous vs. ended gesture) are summarised in Table 24 and visualised in Figure 39.

	condition – gesture	
response	continuous gesture	ended gesture
IPFV	288 (0.62)	93 (0.20)
PFV	180 (0.38)	376 (0.80)

Table 24: *Response types in two experimental conditions*

In both conditions, the majority of responses were congruent with the gesture type. A PFV choice followed after presentation of an ended gesture in 80% of responses. In the continuous gesture condition, 62% of responses were IPFV.

Table 25 shows how the aspectual choice interacts with directedness. In the continuous condition, the proportion of congruent responses was greater when the predicate of the IPFV sentence was undirected. In the ended condition, the directedness difference led to a decreased number of congruent responses compared to stimuli where both sentences involved directed predicates. Note, however, that the number undirected-IPFV stimuli was low (three sentence pairs, 119 responses (1 response was discarded)) and the majority of such trials were presented in the continuous condition.

To investigate the relative contribution of the factors in question to participants' responses, one can take advantage of the hierarchical (mixed-effect) regression models. In this case, the dependent variable is categorical and binary (choice between IPFV and PFV sentence (variable code = `response`) and thus the appropriate method is logistic regression. The factors under scrutiny are gesture type presented in the stim-

[111] At this point, "aspectual choice" is understood formally as a choice of a stimulus sentence which involved an aspectual difference, without presupposing any kind of participants' strategy – that will be discussed below.

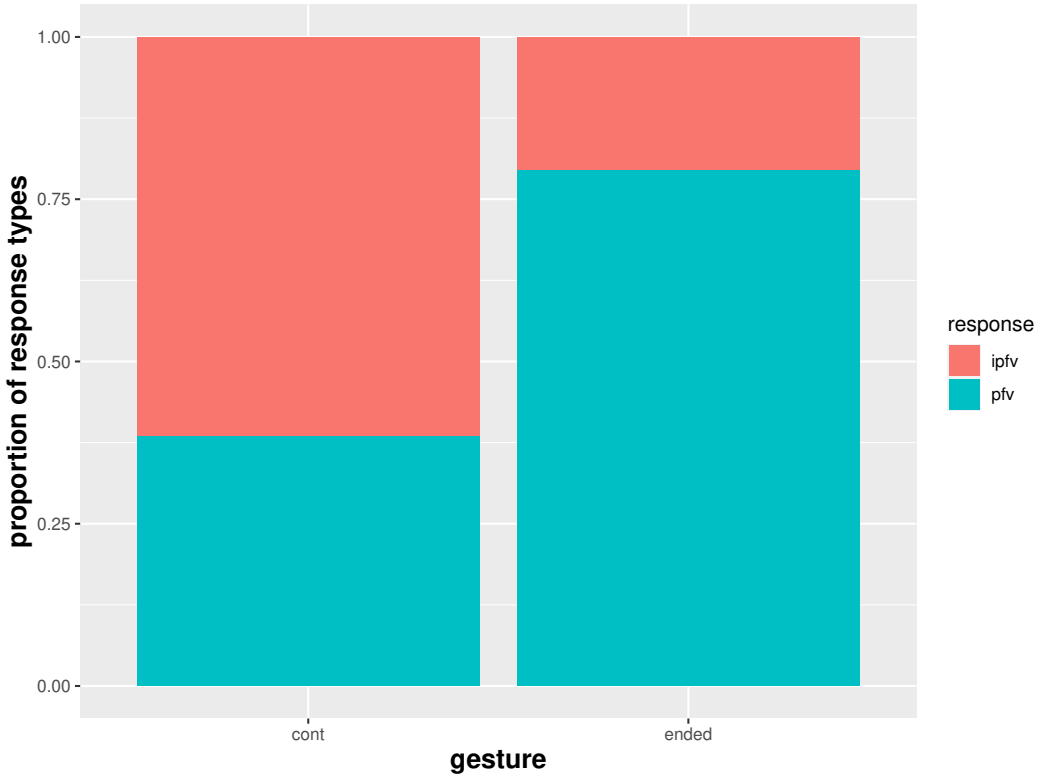


Figure 39: *Relative proportions of response types in two conditions*

	condition – gesture * directedness			
response	cont. (IPFV = undir)	cont. (IPFV = dir)	ended (IPFV = undir)	ended (IPFV = dir)
IPFV	57 (0.71)	231 (0.60)	13 (0.33)	80 (0.19)
PFV	23 (0.29)	157 (0.40)	26 (0.69)	350 (0.81)

Table 25: *Response types in two conditions in interaction with directedness of the IPFV sentence*

ulus videos (`gesture`), and the predictors selected on the basis of the results of the corpus study (predictors with the highest importance according to the random forest analysis). These include directedness of the IPFV sentence (`dir`), object number (`object_num`) and deixis (`deixis`) – these three factors will be investigated in interaction with the gesture type.

Random effects will also be included into the model structure, taking into account inter-participant differences (varying intercepts and slopes for `subject`) as well as variation by individual stimuli (varying intercepts for `item`).

The model was fit using the `glmer()` function from the *lme4* package (Bates et al., 2015) in R.[112] The outcome of the model is summarised in Table 26.

The type of gesture is the only significant predictor of aspectual choice (logit coefficient 2.545, SE = 0.552, z = 4.608, p <0.001). By converting the logit coefficients

[112]Model formula in R syntax: `response ~ gesture + dir * gesture + deixis * gesture + object_num * gesture + (1 + subject|gesture) + (1|item)`

	estimate	SE	*z*	*p*
(*Intercept*)	-0.433	0.331	-1.309	0.191
`gesture:ended`	2.545	0.552	4.608	<0.001
`dir:undir`	-0.129	0.362	-0.356	0.722
`object_num:sg`	-0.190	0.307	-0.619	0.536
`gesture:ended * dir:undir`	-0.411	0.770	-0.534	0.593
`gesture:ended * deixis:yes`	-0.409	0.518	-0.789	0.430
`gesture:ended * object_num:sg`	-0.020	0.485	-0.041	0.967

Table 26: *Summary of the logistic regression model (fixed effects)*

to probabilities,[113] we can see that the model predicts an 89% chance of a PFV response after presentation of an *e*-gesture, compared to a 39% probability of a PFV response after a *c*-gesture (intercept). Such prediction is very close to the actual data (see Table 24).

One way to evaluate the model fit is comparing it by means of a *Likelihood Ratio Test* to a reduced model without the significant predictor. The model comparison performed in R using the function `anova()` revealed a significant difference between the log likelihoods (full model = -514.28, reduced model = -527.70, $\chi^2_{(4)} = 26.841$, $p < 0.001$). The full model also performed better in terms of the degree to which the variance of the responses is explained by the predictors, expressed by R^2 statistic (as a proportion): full model R^2 (marginal) = 0.24 , R^2 (conditional) = 0.44; reduced model R^2 (marginal) = 0.12, R^2 (conditional) = 0.43.[114]

Reaction times

Apart from aspectual choice, the experimental study was also focused on participants reaction times (RT). The key question addressed here is whether there was a tendency to slower responses when the stimuli involved an incongruency between linguistic and gestural encoding of eventuality. Previous research (see especially Kelly et al., 2010, and further evidence under the flag of the Integrated Systems Hypothesis) shown that incongruity between linguistic and gestural information hinders comprehension – which may be manifested by RT latencies.

Given that the stimulus sentences varied in length (although they were relatively well-balanced in terms of syntactic complexity), the RT measures necessarily reflect the varying reading time required by each stimulus item. Therefore, we need to tease apart the response latencies related to the variables in question from the suspected reading time effect.

Before checking the effect of the stimulus length, the dataset was explored in

[113] $P = \frac{exp(log\_odds)}{1-exp(log\_odds)}$

[114] Calculated in R using the `r.squaredGLLM()` function from the *MuMIN* package (Bartoń, 2020). Marginal R^2 corresponds to the proportion of variance described by fixed effects, conditional R^2 considers both fixed and random effects (Winter, 2019, p. 264).

order to find the unwanted variance related to random noise. The overall variance observed in the data was quite large (mean RT = 4332 ms, standard deviation = 2521 ms).

Visualization of the data distribution (Figure 40) reveals the extreme observations that caused the excessive variance.

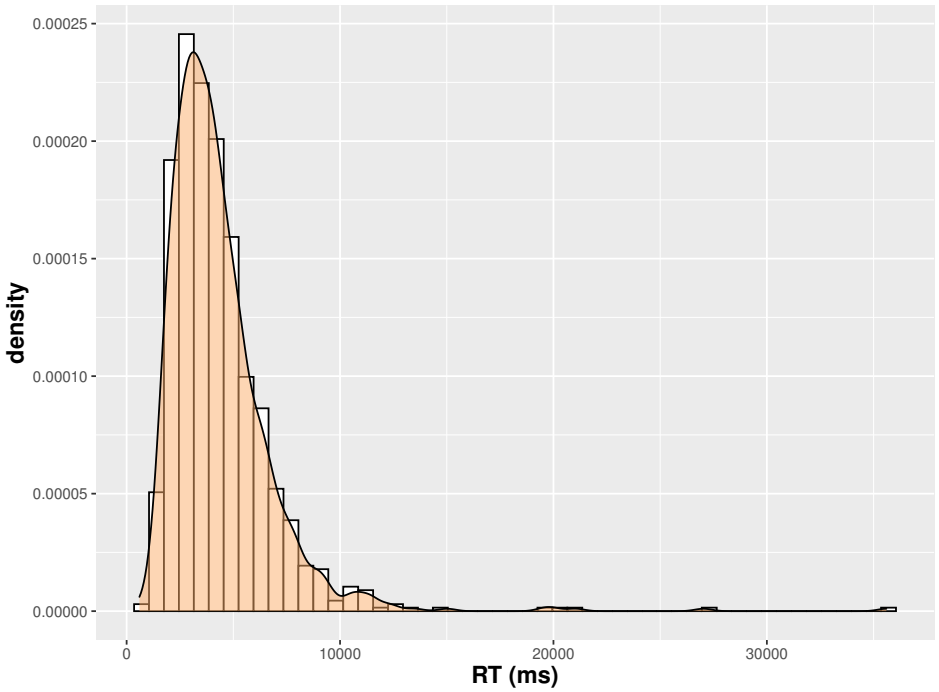


Figure 40: *Histogram – RT distribution before data reduction*

For exploratory studies like this, a rather conservative method of the outlier detection is preferred that would not lead to a data reduction too radical. One of the suitable approaches is the *Median Absolute Deviation* method (Leys et al., 2013) that estimates the outliers as values beyond the range delimited by the median value +/- 2.5 standard deviations. In the case of the present study, only the upper limit makes sense (10 141 ms) as the lower limit is below zero. In total, 23 observations were discarded (2.40%) from the dataset for the subsequent analysis. After data reduction, the mean RT dropped to 4085 ms and the standard deviation was markedly reduced to 1740 ms (see Figure 41 for the RT distribution after the outlier reduction).

In the following step, the effect of reading time was inspected. For each item, length was calculated as the number of characters (without whitespaces) in the item string comprising the two sentences marked by a prefix (A: and B:, respectively) separated by an em dash (—).

A simple correlation analysis (`cor.test()` function, Pearson method) revealed a mild but significant correlation between reaction time (RT) and item length (Pearson's r = 0.20, $t_{(935)}$ = 6.319, p <0.001 (one-tailed)) – see Figure 42.

By fitting a linear regression model (`lm()` function from *lme4* package) we can

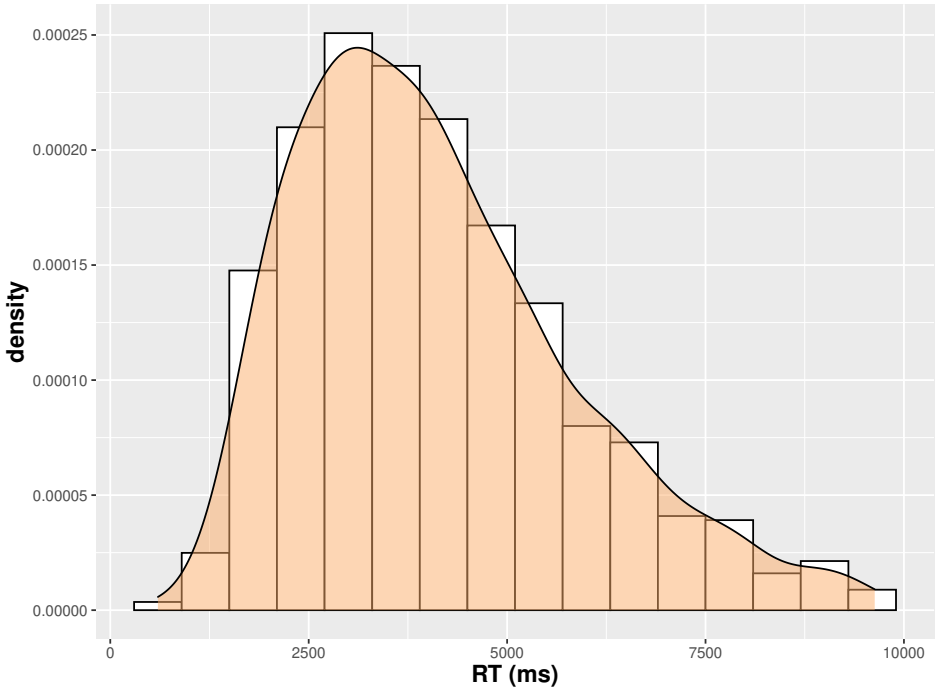


Figure 41: *Histogram – RT distribution after data reduction*

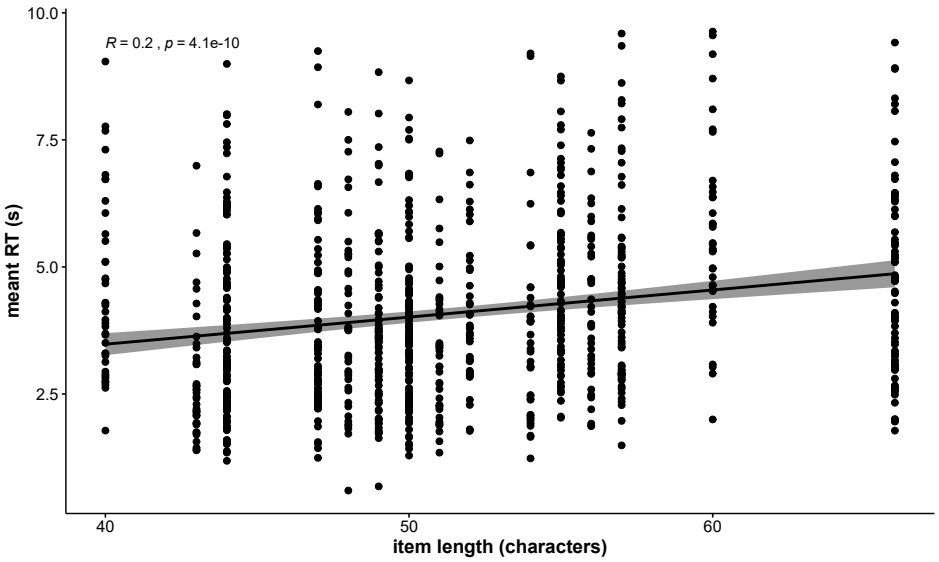


Figure 42: *Scatterplot: correlation between RT and item length*
Grey area around the regression line indicates 95% confidence interval

estimate the average increase of RT when the length increases by one character:[115]

Adding one unit to the parameter length, i.e. increasing the number of characters in a stimulus item by one, leads to average increase of RT by 53 ms (SE = 8 ms) (see Table 27 for the model summary).

The average length-related latency thus may serve as a point of departure for the

[115] Model formula: `rt ~ length`.

	estimate	SE	*t*	*p*
intercept	1.346	0.437	3.081	
`length`	0.053	0.008	6.319	<0.001

Table 27: *Summary of the linear regression model*

subsequent data transformation. The shortest character string (40 characters) serves as the baseline, RTs of the longer strings are reduced in the following manner. For each item, a transformation coefficient was calculated (number of characters above the baseline multiplied by the magnitude of the length effect yielded by the regression model)[116] which was then subtracted from RT. The new variable is referred to as *transformed RT* (tRT).

The visualisation below (Figure 43) shows the distribution of the transformed reaction times across subjects and items, which apparently is considerable (particularly the inter-subject variation) and has to be taken into account in the statistical analysis.

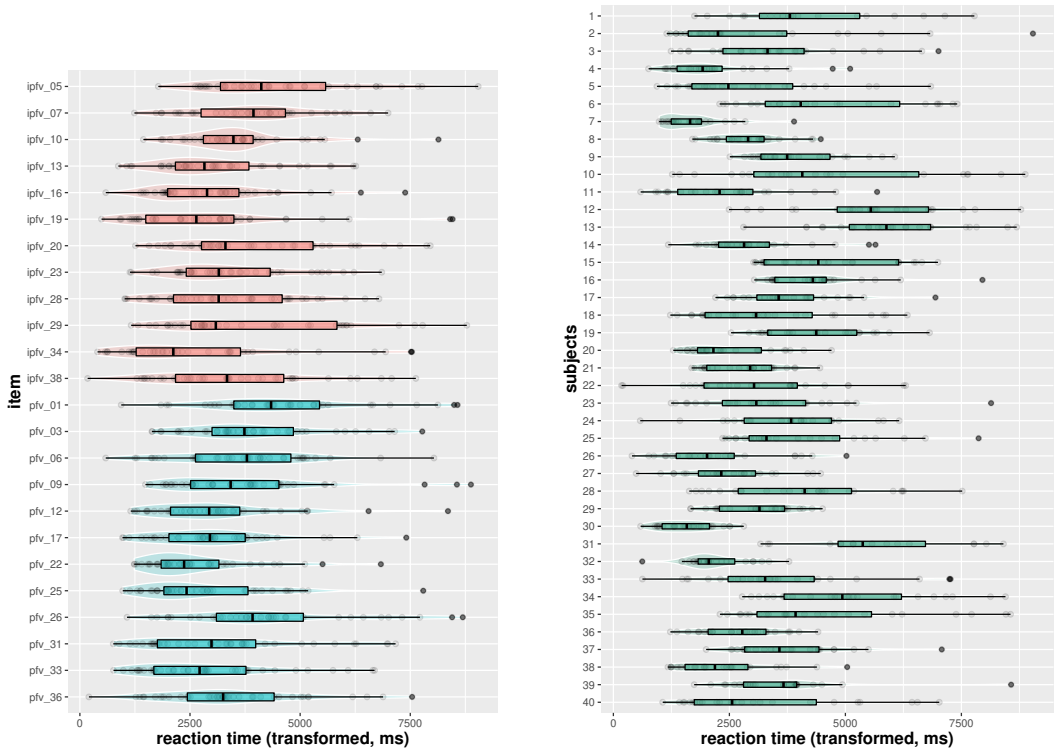


Figure 43: *Transformed reaction times across subjects and items*
Grey dots = individual observations, boxes = interquartile range (middle vertical bar = median)

Another factor related to RT is frequency. As the argument structure follows the same pattern throughout the stimulus items, it is plausible to assume that there might be frequency effect at the lexical level. It has been shown that in sentence comprehension,

[116] Coefficient = $(RT - 40) \times 53$

frequency (and conceptually related measures such as familiarity) of individual words play a pivotal role (see Diessel, 2016, for a review). Sentence comprehension is an incremental process, in which the *focal* part of sentence was reported to be more closely attended during sentence processing (Cutler and Fodor, 1979; Yang et al., 2019). In this study, the critical items differed only in the final segment – the verb – which also marked the sentence focus (by its sentence-final position). Thus, I will only focus on frequency effects related to the verbs.

Table 22 above, provides the relative frequencies for the stimulus verbs retrieved from the Czech National Corpus (Syn2015), measured as the number of instances per million (ipm). Recall that the majority of the verbs occurs more frequently in the PFV form, although the magnitude of the difference varies across the verbs (see the PFV-IPFV ipm difference in Table 22). The PFV-IPFV ipm difference was introduced as an additional variable, to assess the extent to which the choice of the particular aspectual form could have been facilitated by its frequency in usage.

The average tRT across the gestural conditions and response types are summarised in Table 28 and visualised in Figure 44 together with the underlying distribution.

	response	
gesture	IPFV	PFV
ended	3941 (1756)	3340 (1669)
continuous	3453 (1746)	3596 (1620)

Table 28: *Mean tRT*
Measured in milliseconds, standard deviations in parentheses

The fastest RT were recorded when the responses were congruent. On average, the subjects were choosing PFV sentence in the *e*-gesture condition the fastest (3340 ms), IPFV responses in the *c*-gesture condition were slightly slower (3453 ms). The highest average latency (3941 ms) was observed with the least frequent response type: an IPFV sentence chosen after presentation of an *e*-gesture.

The response variable is in this case continuous (`tRT`), and thus an appropriate analytical method is a generalized linear mixed-effect model. The model[117] was fit in R using the `lmer()` function from the *lme4* package, with the following predictors (fixed effects): response type (`response`, PFV vs. IPFV sentence), gesture type (`gesture`), directedness of the IPFV sentence (`dir`), object number (`object_num`), deixis (`deixis`) and the difference between PFV and IPFV ipm (`ipm_diff`). The varying intercepts for subject and item were included as random effects. Table 29 summarizes the model output.

[117] Model formula: `model tRT ~ gesture + response + gesture*response + dir*gesture + object_num*gesture + deixis*gesture + ipm_ratio*response + (1|subject) + (1|item)`

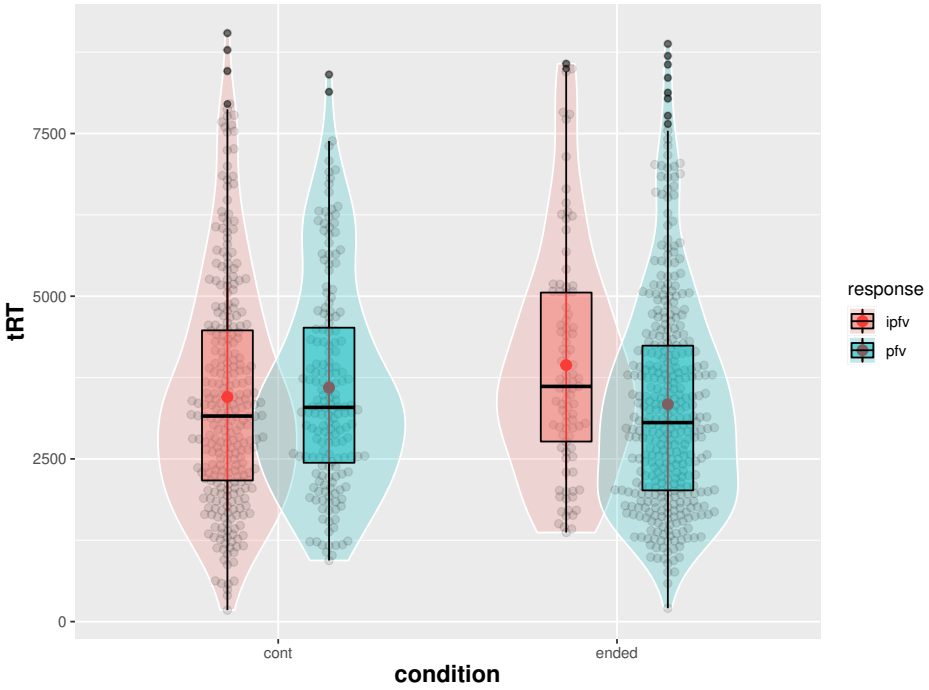


Figure 44: *Mean tRT*
Measured in milliseconds, boxes = interquartile range, middle bar = median value, red dot = mean

	estimate	**SE**	**df**	***t***	***p***
(*Intercept*)	3405.85	285.10	90.84	11.95	
`gesture:ended`	431.18	373.53	60.09	1.15	0.253
`response:pfv`	129.35	131.52	892.24	0.98	0.326
`dir:undir`	504.43	265.87	216.37	1.90	0.059
`object_num:sg`	-139.73	192.02	409.53	-0.73	0.467
`deixis:yes`	65.64	224.89	289.36	0.29	0.771
`ipm_ratio`	18.62	32.75	183.67	0.57	0.570
`gesture:ended * response:pfv`	-454.38	212.93	898.28	-2.13	0.033
`gesture:ended * dir:undir`	-854.83	446.60	301.99	-1.91	0.057
`gesture:ended * object_num:sg`	270.45	301.42	356.32	0.90	0.370
`gesture:ended * deixis:yes`	-268.70	301.63	264.68	-0.89	0.374
`gesture:ended * ipm_ratio`	-35.18	50.68	215.47	-0.69	0.488

Table 29: *Summary of the generalized linear regression model*

From all the predictor variables, including their interactions, the only significant factor was the gesture type in interaction with the response type $\chi^2_{(1)} = 4.93, p = 0.026$).[118] As simple effects, gesture and response type are not significant. According to a model with fixed effects only, directedness is also a reliable predictor, however the inclusion of by-participant and by-item random intercepts was a legitimate step: the mixed-

[118] Calculated using the function `anova()` to compare the full and reduced model (without the significant interaction term), Full model: log-lik = -8109.9, R^2 marginal = 0.45, Reduced model: log-lik = -8112.4, R^2 marginal = 0.45.

effect model is a significantly better fit compared to the fixed-effect-only model ($\chi^2_{(3)}$ = 349.63, p <0.001).[119]

D' measures

Huang and Ferreira (2020) suggested using *Signal Detection Theory* (SDT, see Section 6.1) to analyse linguistic data based on the naïve intuitions. Such approach brings forward a different perspective on the distribution of participants' reactions with focus on their sensitivity to experimental conditions (i.e. to what degree their responses pattern along the stimulus groups) and their bias towards a certain type of response (i.e. a tendency to select the same option regardless of condition).

In the present study, sensitivity (*d'*) is conceptualized as the distance between the distribution of the two response types and bias is, following the procedure from Huang and Ferreira (2020), measured in terms of selection criterion (*c*), which reflects the ratio between number of IPFV responses after an ended gesture and the number of PFV responses after a continuous gesture. Both measures were calculated using the `dprime()` function from the *psycho* package (Makowski, 2018) in R. Sensitivity index was calculated for each participant and the by-participant mean *d'* = 1.34, with 95% confidence interval between 1.01–1.67, this value is significantly different from zero ($t_{(33)}$ = 8.63, p <0.001, one-sample *t*-test, testing the null hypothesis that *d'* equals to zero). In other words, we can say that the participants perceived the gestural conditions as different. Turning to bias, by-participant mean *c* value was -0.29, with 95% confidence interval between -0.42 and -0.15. The results of a one-sample *t*-test show that the *c*-values are significantly different from 0 ($t_{(31)}$ = -4.43, p <0.001) which means that the subjects had a significant tendency towards choosing PFV sentences.

6.2.3 Discussion

Participants' choice between the two stimulus sentences was driven by outer boundedness of the accompanying gestures. When the presented gesture was ended, subjects chose the PFV sentence in 80% of cases. In the continuous gesture condition, 68% of responses favoured a IPFV sentence. The ended-IPFV responses were also characterised by the fastest RT.

Unlike the corpus analysis, the experiment did not reveal an effect of predicate directedness. However, no definitive conclusion can be drawn from this observation, as the undirected predicates, expected to prompt IPFV responses in the continuous condition, were underrepresented in the stimulus material. When the variation between participants and between items was taken into account, the tendencies that were observed in the data proved to be insignificant. The effect of directed remained significant in a model with random slopes and intercepts for subjects, while omitting the

[119] Full model: log-lik = -8109.9, R^2 = 0.45, Reduced model: log-lik = -8284.8, R^2 = 0.34.

random effect of by-item variation. It is reasonable to assume that the model penalized the small number of undirected cases and the question thus lends itself whether, with a more balanced sample, the importance of the effect of directedness might show.

The absence of the effect of frequency in the linear model (tRT) was not surprising as the relative frequency differences between the aspectual forms were not high enough to influence the participants' reaction. The items with a higher-ipm IPFV verb did not influence participants' RT – no frequency effect appeared even when the by-item random intercepts were removed.

Almost 60% of all responses were represented by PFV choices, the bias towards PFV was also reflected in the value of the selection criterion. While the prevalence of the PFV verbs in the list of the most frequent transitive verbs certainly deserves further investigation, another issue needs to be addressed that may be linked to the overall preference for PFV choices. Specifically, there is a possibility that the PFV sentences were, for some reason, simply judged as more acceptable. However, there are some arguments against this hypothesis. First, the participants were not instructed to judge acceptability of the sentences, but to choose "which sentence belonged to the video" and their focus on the visual cues was directed by the filler items, where the gesture exhibited a clearly iconic mapping with one of the verbs. Second, should the participants have used two different strategies for the distractors (a strategy based on gesture) and critical items (a strategy based on the sentence naturalness), it would be likely reflected in RT. That was not the case, as the average tRT for filler items did not differ from the average tRT for the critical items (filler = 3480 ms (SD =1681) vs. critical = 3483 ms (SD = 1704)).

Thus, the results of the statistical analysis suggest that in the continuous gesture condition, the participants' show tendency to associate gesture form with aspect was weaker than in the ended condition. If we inspect the proportions of response times in the particular items (Figure 45), we can see that three sentence pairs in the continuous conditions had a majority (or a half) of incongruent responses Let us first take a closer look at the items in continuous condition with the highest proportion of congruent responses in the *continuous* condition.

[ipfv_29] (85% IPFV): PFV: *Aritmetiku jsem dobře pochopil.*— IPFV: *Aritmetiku jsem dobře chápal.*
('I had a good grasp of arithmetic / I understood arithmetic')

[ipfv_28] (75% IPFV): PFV: *Svého psa doma nechala.*— IPFV: *Svého psa doma nechávala.*
('She left/was leaving her dog at home')

[ipfv_13] (72% IPFV): PFV: *V katalogu si něco vybral.*— IPFV: *V katalogu si něco vybíral.*
('He chose/was choosing something from the catalogue')

The two items that attracted the most IPFV responses [ipfv_29 and ipfv_28] happened to be the sentence pairs that involved a difference in directedness. That supports the assumption that directedness is associated with gesture form after all (although the results of statistical analysis did not support this assumption). More specifically, it seems

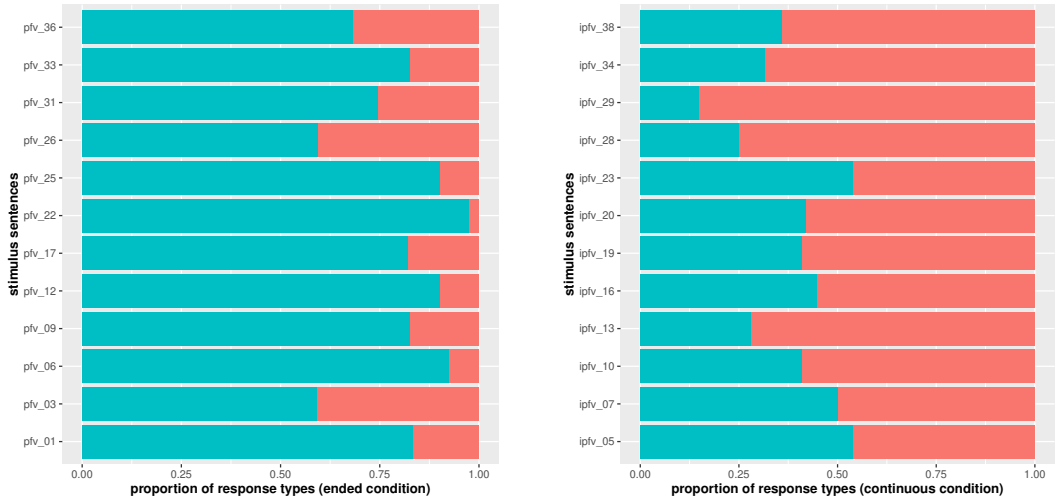


Figure 45: *Proportions of response types by item*
Left = ended condition, right = continuous condition, blue = PFV response, red = IPFV response

that the undirected predicates (or STATE) predicates in IPFV have the strongest association with continuous gestures. Although both sentences in the itemipfv_13 were coded as directed, an undirected construal of the IPFV sentence is possible: it cannot be ruled out that for some of the participants, the activity of *choosing* does not imply a boundary on the *q*-axis: such a construal might have been triggered by the contrastive setting of the two sentences – in which case the participants did not focus the potential endpoint of the IPFV event.

The three items from the continuous condition with the highest proportion of incongruent responses were the following:

[ipfv_23] (54% PFV): PFV: *Mistrovství v šachu začalo.*— IPFV: *Mistrovství v šachu začínalo.*
('The chess championship started/was starting')

[ipfv_05] (54% PFV): PFV: *Ty obědy se vydají.*— IPFV: *Ty obědy se budou vydávat.*
('The meals will be dispensed')

[ipfv_07] (50% PFV): PFV: *Cestu k uznání si nějak našel.*— IPFV: *Cestu k uznání si nějak nacházel.*
('He won/was winning recognition')

In these sentence pairs, both predicates are directed and the IPFV variants of the sentences hardly permit an undirected construal. Otherwise, no apparent explanation for the lack of multimodal effect on the aspectual choice presents itself. In the case of the item [ipfv_07], the PFV responses could have been facilitated by the presence of the vague modifier *nějak* ('somehow'). It is possible that some subjects associated the unbounded and "fluid' hand movement with the vague modifier itself or with the non-specific construal of the event as such, rather than with the specific aspectuality of the verb and thus did not reflect the aspectual difference in their choices, or simply responded randomly.

The results of the experimental study should by no means be considered a binding revelation about the nature of the multimodal construal of eventuality in Czech. The rationale behind the experimental study was to shed light on the findings of the corpus part of this study from a different angle and to point to a reasonable direction of further research which must necessarily follow, building upon not only the evidence presented here, but also the theoretical and methodological issues that came into sight in the process.

The future research should address or avoid the frequency imbalance between PFV and IPFV lemmata. The issue of underrepresentation of undirected predicates in the sample could be resolved by expanding the number of stimuli, which would also enable inclusion of other additional variables.

An evident limitation of the experimental design is the absence of sound in the stimulus material. In the study of multimodal constructions, the acoustic information cannot be disregarded. There is a strong evidence for cross-modal effects in language comprehension (see above, in particular the Integrated Systems Hypothesis): before any generalization about the comprehension of multimodal construals of eventuality can be made, the experiment needs to be re-run in a modified form that would allow for integrating the spoken parts of the utterances in question.

Another limitation of the present study concerns the fact that the same gestures were repeatedly presented in critical stimuli, which could have had an impact on how natural the stimuli looked to the participants.

Despite these limitations, the experimental study provided valuable evidence that, together with the results of the corpus study, adds to the conclusion that in Czech, *e*-gestures and the PFV predicates constitute multimodal eventuality expressions in a relatively stable manner. Importantly, the experiment brings forth support for the ISH hypothesis *beyond* the domain of comprehension of *concrete* meanings. The results of the present study may be interpreted as the evidence for a semantic integration between gestures and abstract verbs that is facilitated via an embodied metaphorical mapping. A follow-up modification of the experiment that would allow for capturing the participants' ERP signatures during the task would help to cast light on the character of the processing of the directly iconic multimodal expressions (such as the filler items in the present study) and multimodal eventuality expressions based on the embodied schemata.

7. Conclusion

> *"Mittler zwischen Hirn und Händen muss das Herz sein!"* (Fritz Lang, Metropolis, 1927)

The primary objective of this study was to explore the multimodal expression of eventuality from a cognitive perspective, building upon a number of studies dedicated to this topic, but departing from them in four major aspects:

(i) focus on gesture production "in the wild" rather than during narratives in controlled settings;

(ii) complex approach to gesture form based on complementary sets of features;

(iii) analysing the data using multifactorial methods and usage-based factors;

(iv) embedding the analysis into the Multimodal CxG framework.

Defining eventuality as a broad semantic domain encompassing the spatial, temporal and force-dynamic aspects of event structure, I focused specifically on the manner of an event's unfolding in a temporal (*aspectuality*) and qualitative-state (*telicity*) dimension, paying particular attention on the construal of event BOUNDEDNESS.

The findings from the corpus study point towards language-specific patterns. In English, the ended gestures tend to co-occur most frequently with the ACH predicates, while continuous gestures are significantly associated with the progressive predicates. In Czech, on the other hand, *e*-gestures are generally associated with directed and non-incremental predicates. Gestural complexity and outer boundedness were found to be correlated: majority of *c*-gestures occurred in the complex forms, whereas *e*-gestures were typically produced in the simplex form. Thus, it is possible to associate directedness with *e*-gestures and incrementality with complex gestures in Czech. In English, incrementality as well as PRG marking increases the chance of observing complex gestural forms.

Two different typological motivations suggest itself for the crosslinguistic divergencies revealed in the present study. First is the overt morphological marking of aspect in Czech, which may be linked to a greater role of gesture in the construal of finer-grained lexical-semantic features such as directedness and incrementality. In English, gesture takes part in ACH/ACT discrimination (in Czech it is marked morphologically by PFV/IPFV distinction).

Second, both in English and Czech, specific clusters of gestural forms appear to be associated with different lexicalization patterns at the morphological level. This is the case of English PRG constructions and Czech PFV.

The two tendencies may seem contradictory (especially in Czech) however it may very well be the case that they both are in place: the first as a visible manifestation of profiling operations, the second as a sign of multimodal constructionalization. The effects and interrelations between the two tendencies represent key issues for the future research.

In general terms, the findings of this study add up to the accumulated evidence supporting the Interface Hypothesis. Unlike the bulk of the studies carried out under the umbrella of the Interface Hypothesis and related theoretical models, such as *thinking and gesturing for speaking*, the present study did not rely upon clearly iconic gestural representations of various aspects of motion events. Instead, it took into consideration the entire range of gestural forms that co-occurred with predicates, focusing on the realization of the phonological features of ending accentuation and acceleration. Apart from end-marking, also the inner complexity of the gestural movement was in focus. Thus, a greater variety of language-specific multimodal expression patterns could have been identified. A one-dimensional view (bounded vs. unbounded gesture) would allow for capturing only one side of the coin, ignoring the association between inner complexity and incremental aspectual types. Moreover, it would obscure one of the critical findings concerning the complementarity between the outer- and inner-boundedness features.

The results of the experiment, focusing on perception of multimodal eventuality expressions in Czech, albeit limited, are in line with the Integrated Systems Hypothesis, providing evidence for the gestural discrimination effect at a more abstract level. However, to address the role of gesture in the processing at the level of grammatical constructions, a follow-up experiment is required to factor in the missing piece of the integrated system: the role of prosody and phonetic processing.

While the experiment clearly demonstrated that the association between PFV and *e*-gestures is also manifested during the comprehension processes, the results of the mixed-effect regression models did not confirm the effect of directedness. Nevertheless, a closer inspection of the data suggested that the possible effect directedness cannot be ruled out and needs to be revisited in a follow-up study, taking a different methodological approach that would allow for teasing apart the factors of aspect and directedness.

*
* *

For the study of multimodal constructions, it is crucial to predict which variance is and is not relevant for the conventionalization of a multimodal pattern. First, such constructional approach must be embraced that takes into account all relevant features

and constraints from all linguistic domains, including discourse functions, potential register and genre influence and so on. Second, an onomasiological approach should be adopted, especially in cross-linguistic studies. Such approach is based on the delimitation of the properties of the conceptual frames prior to tracing down the specifics of their construals. If approached this way, predictions can be made about potential prominence and profiling in the target linguistic structures. To attest a pattern as conventionalizing or conventionalized, the observed variance in the target constructions must prove to be significantly reduced within the profiled elements.

The results of this study show that this may be the case for multimodal marking of aspectual types: the most stable patterns emerge where the most prominent feature of the internal structure of an aspectual type is highlighted in gesture. Moreover, it has been shown for constructions in general that some type of variation does not undermine the status of the construction itself, on the contrary, it can signify the productivity of the construction as some of its elements can be varied and thus adjusted to new contexts and communicative needs without losing the core meaning and function. Moreover, the results also point in the direction that has been discussed within CxG for some time, i.e. that speakers probably generalize and entrench constructions on lower level than originally expected, thus forming more complex patterns (such as for particular TAM combinations and/or aspectual (sub)types) rather than general patterns for PFV or IPFV constructions as such. Following this direction, the future steps should be heading towards bringing more evidence of conventionalizing aspectual patterns in the languages in focus, such as incremental events in combination with continuous and complex gestures in both languages, directed PFV and IPFV types in Czech and contrastive patterns of ACH and ACC in English.

A number of paths open up for the future studies. A follow-up ERP study using expanded and improved design of the present experiment, suggested in Chapter 6, is one of them. Currently, a new multimodal corpus of Czech interactions[120] is in development, that will serve as a suitable source of the data for follow-up corpus-based studies. Among other topics, these studies will focus on the multifunctionality of *e*-gestures, in order to tease apart their association with prosodic emphasis and information structure on the one hand, and eventuality on the other. The corpus will also prepare the ground for subsequent experimental studies, in which the corpus material will be utilised in the design of experimental stimuli.

[120] https://sites.google.com/view/epocc/czico

References

Abner, N., K. Cooperrider, and S. Goldin-Meadow

2015. Gesture for Linguists: A Handy Primer: Gesture for Linguists. *Language and Linguistics Compass*, 9(11):437–451.

Albert, S. and J. P. De Ruiter

2018. Improving Human Interaction Research through Ecological Grounding. *Collabra: Psychology*, 4(1):24.

Alibali, M. W., A. Yeo, A. B. Hostetter, and S. Kita

2017. Representational gestures help speakers package information for speaking. In *Why Gesture? How the Hands Fuction in Speaking, Thinking and Communicating*, R. B. Church, M. W. Alibali, and S. D. Kelly, eds., Pp. 15–37. Amsterdam: John Benjamins.

Allen, S., A. Özyürek, S. Kita, A. Brown, R. Furman, T. Ishizuka, and M. Fujii

2007. Language-specific and universal influences in children's syntactic packaging of Manner and Path: A comparison of English, Japanese, and Turkish. *Cognition*, 102(1):16–48.

Andrén, M.

2010. *Children's Gestures from 18 to 30 Months*. Doctoral dissertation, Lund University, Lund.

Arbib, M. A.

2013. Précis of How the brain got language: The Mirror System Hypothesis. *Language and Cognition*, 5(2-3):107–131.

Arbib, M. A., B. Gasser, and V. Barrès

2014. Language is handy but is it embodied? *Neuropsychologia*, 55:57–70.

Argyle, M.

1975. *Bodily Communication.* New York, NY: International Universities Press.

Arik, E.

2012. Space, time, and iconicity in Turkish Sign Language (TID). *Trames. Journal of the Humanities and Social Sciences*, 16(4):345.

Athanasopoulos, P. and E. Bylund

2013. Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science*, 37(2):286–309.

Austin, J. L.

1962. *How to do things with words: [the William James lectures delivered at Harvard University in 1955]*. Oxford: Clarendon Press.

Bach, E.
1986. The algebra of events. *Linguistics and Philosophy*, 9:5–16.

Baldassano, C., J. Chen, A. Zadbood, J. W. Pillow, U. Hasson, and K. A. Norman
2017. Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, 95(3):709–721.e5.

Baragwanath, N.
2007. Anna Bahr-Mildenburg, gesture, and the Bayreuth style. *The Musical Times*, 148(1901):63–74.

Barsalou, L. W.
1999. Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(04):637–660.

Barsalou, L. W. and K. Wiemer-Hastings
2005. Situating abstract concepts. In *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, D. Pecher and R. A. Zwaan, eds., Pp. 129–163. Cambridge: Cambridge University Press.

Bartoń, K.
2020. MuMIn: Multi-model inference. https://cran.r-project.org/package=MuMIn.

Bates, D., M. Mächler, B. Bolker, and S. Walker
2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1).

Baynton, D. C.
1996. *Forbidden signs: American culture and the campaign against sign language*. Chicago, IL: University of Chicago Press.

Beattie, G. and H. Shovelton
1999. Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, 123(1-2):1–30.

Beattie, G. and H. Shovelton
2002. What properties of talk are associated with the generation of spontaneous iconic hand gestures? *British Journal of Social Psychology*, 41(3):403–417.

Becker, R., A. Cienki, A. Bennett, C. Cudina, C. Debras, Z. Fleischer, M. Haaheim, T. Müller, K. Stec, and A. Zarcone
2011. Aktionsarten in Speech and Gesture. *GESPIN: Gesture and Speech in Interaction Proceedings*, Pp. 30–35.

Becker, R. and M. Gonzalez Marquez
2018. Comprehension of event construal from multimodal communication. In *Aspectuality across Languages*, A. J. Cienki and O. K. Iriskhanova, eds., Pp. 161–178. Amsterdam: John Benjamins.

Bendor, D. and X. Wang
2006. Cortical representations of pitch in monkeys and humans. *Current Opinion in Neurobiology*, 16(4):391–399.

Bergs, A. and G. Diewald, eds.
2009. *Contexts and Constructions*. Amsterdam: John Benjamins.

Bernardis, P., E. Salillas, and N. Caramelli
2008. Behavioural and neurophysiological evidence of semantic interaction between iconic gestures and words. *Cognitive Neuropsychology*, 25(7-8):1114–1128.

Bertrand, R., P. Blache, R. Espesser, G. Ferré, C. Meunier, B. Priego-Valverde, and S. Rauzy
2009. The "Corpus of Interactional Data" (CID) - Multimodal annotation of conversational speech" Le CID - Corpus of Interactional Data. Annotation et exploitation multimodale de parole conversationnelle. *Traitement Automatique des Langues*, 49(3):105–134.

Birdwhistell, R. L.
1970. *Kinesics and context: essays on body motion communication*. Philadelphia, PA: University of Pennsylvania Press.

Boas, H. C. and I. A. Sag, eds.
2012. *Sign-based construction grammar*. Stanford, CA: CSLI Publications.

Boersma, P.
2001. Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10):341–345.

Bohle, U.
2014. Contemporary classifications. In *Body - Language - Communication*, C. Müller, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem, eds., Pp. 1453–1461. Berlin: De Gruyter.

Bohr, N.
1928. The Quantum postulate and the recent development of atomic theory. *Nature*, 121:580–590.

Bolinger, D. L.
1983. Gesture and intonation. *American Speech*, 58(2):156–174.

Boroditsky, L.
2001. Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43(1):1–22.

Bosker, H. R. and D. Peeters
2020. Beat gestures influence which speech sounds you hear. preprint, Neuroscience. DOI: 10.1101/2020.07.13.200543.

Boutet, D.
2010. Structuration physiologique de la gestuelle: modèle et tests. *LIDIL*, 42:77–96.

Breiman, L.
2001. Random forests. *Machine Learning*, 45:5–32.

Bressem, J.
2013. A linguistic perspective on the notation of form features in gestures. In *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1*, C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and S. Tessendorf, eds., Pp. 1079–1098. Berlin: De Gruyter.

Bressem, J. and C. Müller
2017. The "Negative-Assessment-Construction" – A multimodal pattern based on a recurrent gesture? *Linguistics Vanguard*, 3(s1).

Browman, C. P. and L. Goldstein
1992. Articulatory phonology: An overview. *Phonetica*, 49(3-4):155–180.

Brône, G. and E. Zima
2014. Towards a dialogic construction grammar: Ad hoc routines and resonance activation. *Cognitive Linguistics*, 25(3):457–495.

Butterworth, G.
2003. Pointing is the royal road to language for babies. In *Pointing: Where Language, Culture, and Cognition Meet*, S. Kita, ed., Pp. 9–33. Mahwah, NJ: Lawrence Erlbaum Associates.

Bylund, E.
2009. Effects of age of L2 acquisition on L1 event conceptualization patterns. *Bilingualism: Language and Cognition*, 12(3):305–322.

Bylund, E., P. Athanasopoulos, and M. Oostendorp
2013. Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics*, 51(5):286–309.

Bögels, S. and S. C. Levinson

2017. The Brain Behind the Response: Insights Into Turn-taking in Conversation From Neuroimaging. *Research on Language and Social Interaction*, 50(1):71–89.

Börstell, C. and R. Östling

2017. Iconic Locations in Swedish Sign Language: Mapping Form to Meaning with Lexical Databases. In *Proceedings of the 21st Nordic Conference on Computational Linguistics (NODALIDA 2017), NEALT Proceedings seriers 29*, J. Tiedemann, ed., Pp. 221–225, Gothenburg. Linköping University Electronic Press.

Bühler, K.

1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: G. Fischer.

Calbris, G.

2003. From cutting an object to a clear cut analysis: Gesture as the representation of a preconceptual schema linking concrete actions to abstract notions. *Gesture*, 3(1):19–46.

Calbris, G.

2008. From left to right...: Coverbal gestures and their symbolic use of space. In *Metaphor and Gesture*, A. J. Cienki and C. Müller, eds., Pp. 27–54. Amsterdam: John Benjamins.

Carletta, J.

2006. Announcing the AMI Meeting Corpus. *The ELRA Newsletter*, 11(1):3–5.

Casasanto, D.

2009. Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, 138(3):351–367.

Casasanto, D.

2011. Different bodies, different minds: The body specificity of language and thought. *Current Directions in Psychological Science*, 20(6):378–383.

Casasanto, D.

2014. Bodily relativity. In *Routledge Handbook of Embodied Cognition*, Pp. 108–117. New York, NY: Routledge.

Casasanto, D.

2016. A shared mechanism of linguistic, cultural, and bodily relativity: Experiential relativity. *Language Learning*, 66(3):714–730.

Casasanto, D. and E. G. Chrysikou

2011. When left is "right": Motor fluency shapes abstract concepts. *Psychological Science*, 22(4):419–422.

Casasanto, D. and K. Jasmin
2010. Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. *PLoS ONE*, 5(7):e11805.

Chomsky, N., R. W. Rieber, and G. Voyat, eds.
1983. *Dialogues on the psychology of language and thought: conversations with Noam Chomsky, Charles Osgood, Jean Piaget, Ulric Neisser, and Marcel Kinsbourne*. New York, NY: Plenum Press.

Christensen, P. and M. Gullberg
2016. Musical pitch metaphors in speech and gesture: evidence from Swedish and Turkish. Presented at Perception Metaphor Workshop, MPI for Psycholinguistics, Nijmegen.

Chromý, J.
2014. Impact of tense on the interpretation of bi-aspectual verbs in czech. *Studie z aplikované lingvistiky / Studies in applied linguistics*, 5(2):87–97.

Church, R. B., M. W. Alibali, and S. D. Kelly, eds.
2017. *Why Gesture? How the Hands Fuction in Speaking, Thinking and Communicating*. Amsterdam: John Benjamins.

Cienki, A. J.
2005. Image schemas in cognitive linguistics. In *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, B. Hampe, ed., Pp. 421–442. Berlin: Mouton de Gruyter.

Cienki, A. J. and O. K. Iriskhanova, eds.
2018. *Aspect across languages: event construal in speech and gesture*. Amsterdam: John Benjamins.

Cienki, A. J. and C. Müller, eds.
2008. *Metaphor and gesture*, number v. 3. Amsterdam: John Benjamins. OCLC: 202548320.

Clowes, R. W. and D. Mendonça
2015. Representation Redux: Is there still a useful role for representation to play in the context of embodied, dynamicist and situated theories of mind? *New Ideas in Psychology*, 40:26–47.

Cochran, W. G.
1954. Some Methods for Strengthening the Common χ 2 Tests. *Biometrics*, 10(4):417.

Cohen, J.
1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.

Comrie, B.
1976. *Aspect*. Cambridge: Cambridge University Press.

Comrie, B., M. Haspelmath, and B. Bickel
2008. The Leipzig glossing rules: Conventions for interlinear morpheme-by-morpheme glosses.

Connolly, J. H.
2010. Accommodating multimodality in Functional Discourse Grammar. *Web Papers in Functional Discourse Grammar*, (83):1–18.

Cooperrider, K., J. Slotta, and R. Núñez
2018. The Preference for Pointing With the Hand Is Not Universal. *Cognitive Science*, 42(4):1375–1390.

Cooperrider, K. A.
2011. *Reference in action : links between pointing and language*. Doctoral dissertation, UC San Diego.

Corballis, M. C.
2002. *From hand to mouth: the origins of language*. Princeton, NJ: Princeton University Press. OCLC: 474436988.

Cormier, K., A. Schembri, and B. Woll
2013. Pronouns and pointing in sign languages. *Lingua*, 137:230–247.

Croft, W.
2001. *Radical construction grammar: syntactic theory in typological perspective*. Oxford ; New York, NY: Oxford University Press.

Croft, W.
2012. *Verbs: aspect and causal structure*. Oxford: Oxford University Press. OCLC: ocn757930806.

Croft, W.
2016. Typology and the future of Cognitive Linguistics. *Cognitive Linguistics*, 27(4):587–602.

Croft, W. and D. A. Cruse
2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.

Croft, W., P. Pešková, and M. Regan
2016. Annotation of causal and aspectual structure of events in RED: a preliminary report. In *4th Events Workshop, 15th Annual Conference of the North American Chapter of the Association of Computational Linguistics: Human Language Technologies (NAACL-HLT 2016)*, Pp. 8–17, Stroudsbourg, PA. Association for Computational Linguistics.

Cutler, A. and J. A. Fodor
1979. Semantic focus and sentence comprehension. *Cognition*, 7(1):49–59.

Dahl, Ö.
1985. *Tense and aspect systems*. Oxford: B. Blackwell.

Dahl, Ö.
1990. Standard Average European as an exotic language Östen Dahl. In *Toward a Typology of European Languages*, J. Bechert, G. Bernini, and C. Buridant, eds., Pp. 3–8. Berlin: De Gruyter.

Daneš, F.
1971. Pokus o strukturní analýzu slovesných významů. *Slovo a slovesnost*, 32:193–207.

de Ruiter, J. P.
2000. The production of gesture and speech. In *Language and Gesture*, D. McNeill, ed., Pp. 248–311. Cambridge: Cambridge University Press.

de Ruiter, J. P.
2017. The asymmetric redundancy of gesture and speech. In *Why Gesture? How the Hands Fuction in Speaking, Thinking and Communicating*, R. B. Church, M. W. Alibali, and S. D. Kelly, eds., Pp. 59–75. Amsterdam: John Benjamins.

de Ruiter, J. P. and S. Albert
2017. An Appeal for a Methodological Fusion of Conversation Analysis and Experimental Psychology. *Research on Language and Social Interaction*, 50(1):1–18.

de Vos, C.
2015. The Kata Kolok Pointing System: Morphemization and Syntactic Integration. *Topics in Cognitive Science*, 7(1):150–168.

di Pellegrino, G., L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti
1992. Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91(1):176–180.

Dickey, S. M.
2000. *Parameters of Slavic aspect: a cognitive approach*. Stanford, CA: Center for the Study of Language and Information.

Diessel, H.
2016. Frequency and lexical specificity in grammar: A critical review. In *Experience Counts: Frequency Effects in Language*, H. Behrens and S. Pfänder, eds., Pp. 209–238. Berlin: De Gruyter.

Dingemanse, M.
2011. *The Meaning and Use of Ideophones in Siwu*. Doctoral dissertation, Radboud Universiteit, Nijmegen.

Dingemanse, M. and N. J. Enfield
2015. Other-initiated repair across languages: towards a typology of conversational structures. *Open Linguistics*, 1(1):98–118.

Dingemanse, M. and S. Floyd
2014. Conversation across cultures. In *The Cambridge Handbook of Linguistic Anthropology*, J. Sidnell and N. Enfield, J., eds., Pp. 447–480. Cambridge: Cambridge University Press.

Dingemanse, M., S. G. Roberts, J. Baranova, J. Blythe, P. Drew, S. Floyd, R. S. Gisladottir, K. H. Kendrick, S. C. Levinson, E. Manrique, G. Rossi, and N. J. Enfield
2015. Universal Principles in the Repair of Communication Problems. *PLOS ONE*, 10(9):e0136100.

Dipper, L., M. Pritchard, G. Morgan, and N. Cocks
2015. The language–gesture connection: Evidence from aphasia. *Clinical Linguistics & Phonetics*, 29(8-10):748–763.

Dipper, S., M. Goetze, and S. Skopeteas, eds.
2007. *Information structure in cross-linguistic corpora: annotation guidelines for phonology, morphology, syntax, semantics and information structure*. Potsdam: Universitätsverlag Potsdam.

Divjak, D.
2011. Predicting aspectual choice in modal constructions: a quest for the Holy Grail? In *Slavic Linguistics in a Cognitive framework*, M. Grygiel and L. A. Janda, eds., Pp. 67–85. Berlin: Peter Lang.

Divjak, D., N. Levshina, and J. Klavan
2016. Cognitive Linguistics: Looking back, looking forward. *Cognitive Linguistics*, 27(4):447–463.

Dolscheid, S., S. Shayan, A. Majid, and D. Casasanto
2013. The Thickness of Musical Pitch: Psychophysical Evidence for Linguistic Relativity. *Psychological Science*, 24(5):613–621.

Dowty, D. R.
1991. Thematic Proto-Roles and Argument Selection. *Language*, 67:547–619.

Drijvers, L., A. Özyürek, and O. Jensen
2018. Hearing and seeing meaning in noise: Alpha, beta, and gamma oscillations

predict gestural enhancement of degraded speech comprehension. *Human Brain Mapping*, P. Advance online publication.

Du Bois, J. W.
2014. Towards a dialogic syntax. *Cognitive Linguistics*, 25(3):359–410.

Duncan, S. D.
2002. Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture*, 2(2):183–206.

Ebert, C., S. Evert, and K. Wilmes
2011. Focus marking via gestures. In *Proceedings of Sinn and Bedeutung 15*, I. Reich, E. Horch, and D. Pauly, eds., Pp. 193–208. Saarbrücken: Saarland University Press.

Efron, D.
1941. *Gesture and Environment*. New York, NY: King's Crown Press.

Eitan, Z. and R. Timmers
2010. Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, 114(3):405–422.

Ekman, P. and W. V. Friesen
1969. The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. *Semiotica*, 1(1):49–98.

Emmorey, K., H. B. Borinstein, R. Thompson, and T. H. Gollan
2008. Bimodal bilingualism. *Bilingualism: Language and Cognition*, 11(1):43–61.

Enfield, N. J.
2009. *The anatomy of meaning: speech, gesture, and composite utterances*, number 8. Cambridge, UK ; New York: Cambridge University Press. OCLC: ocn268793359.

Evans, N. and S. C. Levinson
2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32(05):429–492.

Ferré, G.
2010. Timing Relationships between Speech and Co-Verbal Gestures in Spontaneous French. In *Workshop on Multimodal Corpora*, Pp. 86–91, Malta.

Feyereisen, P.
2006. How could gesture facilitate lexical access? *Advances in Speech Language Pathology*, 8(2):128–133.

Feyereisen, P.
2013. Psycholinguistics of speech and gestures: Production, comprehension, architecture. In *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. (Handbooks of Linguistics and Communication Science 38.1.)*, Pp. 156–168. Berlin: De Gruyter Mouton.

Fibigerová, K. and M. Guidetti
2018. The impact of language on gesture in descriptions of voluntary motion in Czech and French adults and children. *Language, Interaction and Acquisition. Langage, Interaction et Acquisition*, 9(1):101–136.

Filip, H.
1999. *Aspect, eventuality types, and nominal reference*. New York, NY: Garland Pub.

Fillmore, C., P. Kay, and M. O'Connor
1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. *Language*, 64(3):501–538.

Fillmore, C. J.
1982. Frame semantics. In *Linguistics in the Morning Calm*, H. Ŏnŏ Hakhoe, ed., Pp. 111–137. Seoul: Hanshin Publishing Co.

Fillmore, C. J.
2012. Encounters with Language. *Computational Linguistics*, 38(4):701–718.

Fillmore, C. J.
2013. Berkeley Construction Grammar. In *Oxford Handbook of Construction Grammar*, T. Hoffmann and G. Trousdale, eds., volume 1, Pp. 111–132. Oxford: Oxford University Press.

Firbas, J.
1992. *Functional sentence perspective in written and spoken communication*. Cambridge: Cambridge University Press.

Fischer, K.
2010. Beyond the sentence: Constructions, frames and spoken interaction. *Constructions and Frames*, 2(2):185–207.

Fischer, K.
2015. Conversation, Construction Grammar, and cognition. *Language and Cognition*, 7(4):563–588.

Fischer, M. H. and R. A. Zwaan
2008. Embodied Language: A Review of the Role of the Motor System in Language Comprehension. *Quarterly Journal of Experimental Psychology*, 61(6):825–850.

Floyd, S.
2016. Modally hybrid grammar?: Celestial pointing for time-of-day reference in Nheengatú. *Language*, 92(1):31–64.

Fodor, J.
1976. *The Language Of Thought*. New York, NY: Crowell Press.

Frawley, W.
1992. *Linguistic semantics*. Hillsdale, N.J: L. Erlbaum Associates.

Freed, A.
1979. *The Semantics of English Aspectual Complementation - Ghent University Library*. PhD dissertation, Universiteit Gent, Gent.

Fried, M.
2015. Construction Grammar. In *Syntax – Theory and Analysis. An International Handbook*, A. Alexiadou and T. Kiss, eds., Pp. 974–1003. Berlin: Mouton de Gruyter.

Fried, M. and J.-O. Östman
2004. Construction Grammar: A thumbnail sketch. In *Constructional Approaches to Language*, M. Fried and J.-O. Östman, eds., volume 2, Pp. 11–86. Amsterdam: John Benjamins.

Fried, M. and J.-O. Östman
2005. Construction Grammar and spoken language: The case of pragmatic particles. *Journal of Pragmatics*, 37(11):1752–1778.

Fulka, J.
2017. Ruce a svět: Merleau-Ponty, gesto a znakový jazyk. *Studie z aplikované lingvistiky / Studies in applied linguistics*, 8(1):43–52.

Gallagher, S.
2005. *How the body shapes the mind*. Oxford: Clarendon Press. OCLC: ocm56964733.

Gallagher, S. and D. D. Hutto
2008. Understanding others through primary interaction and narrative practice. In *Converging Evidence in Language and Communication Research*, J. Zlatev, T. P. Racine, C. Sinha, and E. Itkonen, eds., volume 12, Pp. 17–38. Amsterdam: John Benjamins.

Gallese, V. and G. Lakoff
2005. The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3-4):455–479.

Gamer, M., J. Lemon, I. Fellows, and P. Singh
2012. irr: Various Coefficients of Interrater Reliability and Agreement. https://cran.r-project.org/web/packages/irr/index.html.

Garfinkel, H.
1967. *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice-Hall.

Geeraerts, D.
2010. *Theories of lexical semantics*. Oxford: Oxford University Press. OCLC: ocn429750193.

Geeraerts, D.
2016. The sociosemiotic commitment. *Cognitive Linguistics*, 27(4):527–542.

Gibbs, R. W.
2006. *Embodiment and cognitive science*. Cambridge ; New York: Cambridge University Press. OCLC: ocm57531469.

Gilhooly, K. J. and R. H. Logie
1980. Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, 12(4):395–427.

Givón, T.
1991. Isomorphism in the Grammatical Code: Cognitive and Biological Considerations. *Studies in Language*, 15(1):85–114.

Goodwin, C.
2000. Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32(10):1489–1522.

Goodwin, C.
2003. Pointing as situated practice. In *Pointing: Where Language, Culture, and Cognition Meet*, S. Kita, ed., Pp. 217–242. Mahwah, NJ: Lawrence Erlbaum Associates.

Goodwin, C.
2013. The co-operative, transformative organization of human action and knowledge. *Journal of Pragmatics*, 46(1):8–23.

Grice, P. H.
1975. Logic and Conversation. In *Syntax and Semantics, Vol. 3, Speech Acts*, P. Cole and J. L. Morgan, eds., Pp. 41–58. New York, NY: Academic Press.

Gries, S. T.
2019. On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. *Corpus Linguistics and Linguistic Theory*. https://www.degruyter.com/view/journals/cllt/ahead-of-print/article-10.1515-cllt-2018-0078/article-10.1515-cllt-2018-0078.xml.

Gumperz, J. J. and S. C. Levinson, eds.
1996. *Rethinking linguistic relativity*. Cambridge; New York, NY: Cambridge University Press.

Haddington, P.
2007. Positioning and alignment as activities of stancetaking in news interviews. In *Pragmatics & Beyond New seriers*, R. Englebretson, ed., volume 164, Pp. 283–317. Amsterdam: John Benjamins.

Halina, M., K. Liebal, and M. Tomasello
2018. The goal of ape pointing. *PLOS ONE*, 13(4):e0195182.

Hallett, M.
2007. Transcranial Magnetic Stimulation: A Primer. *Neuron*, 55(2):187–199.

Harnad, S.
1990. The Symbol Grounding Problem. *Physica D*, (42):335–346.

Harrison, S.
2014. The organisation of kinesic ensembles associated with negation. *Gesture*, 14(2):117–140.

Harré, R. and G. Gillett
1994. *The discursive mind*. Thousand Oaks, CA: Sage Publications.

Haspelmath, M.
2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, 86(3):663–687.

Hassemer, J.
2016. *Towards a theory of Gesture Form Analysis. Imaginary forms as part of gesture conceptualisation, with empirical support from motion-capture data (PhD Thesis)*. PhD dissertation, Humboldt-Universität zu Berlin, Berlin.

Hassemer, J. and L. McCleary
2018. The multidimensionality of pointing. *Gesture*, 17(3):416–461.

Hassemer, J. and B. Winter
2018. Decoding Gestural Iconicity. *Cognitive Science*, 42(8):3034–3049.

Heisenberg, W.
1927. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43:172–198.

Heisenberg, W.
1958. The Representation of Nature in Contemporary Physics. *Daedalus*, 87(3):95–108.

Hewes, G. W.
1973. Primate Communication and the Gestural Origin of Language. *Current Anthropology*, 14:5–24.

Hickmann, M., H. Hendriks, and M. Gullberg
2011. Developmental perspectives on the expression of motion in speech and gesture: A comparison of French and English. *Language, Interaction and Acquisition*, 2(1):129–156.

Hickok, G.
2014. *The myth of mirror neurons: the real neuroscience of communication and cognition*, first edition edition. New York, NY: W. W. Norton & Company.

Hinnell, J.
2018. The multimodal marking of aspect: The case of five periphrastic auxiliary constructions in North American English. *Cognitive Linguistics*, 29(4):773–806.

Holle, H. and T. C. Gunter
2007. The Role of Iconic Gestures in Speech Disambiguation: ERP Evidence. *Journal of Cognitive Neuroscience*, 19(7):1175–1192.

Holle, H., C. Obermeier, M. Schmidt-Kassow, A. D. Friederici, J. Ward, and T. C. Gunter
2012. Gesture Facilitates the Syntactic Analysis of Speech. *Frontiers in Psychology*, 3:74.

Holler, J.
2013. Experimental methods in co-speech gesture research. In *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1*, C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and S. Tessendorf, eds., Pp. 837–856. Berlin, Boston: De Gruyter.

Holler, J. and G. Beattie
2003. Pragmatic aspects of representational gestures. Do speakers use them to clarify verbal ambiguity for the listener? *Gesture*, 3(2):127–154.

Holler, J. and K. H. Kendrick
2015. Unaddressed participants' gaze in multi-person interaction: optimizing recipiency. *Frontiers in Psychology*, 6:98.

Holler, J., K. H. Kendrick, M. Casillas, and S. C. Levinson, eds.
2016. *Turn-Taking in Human Communicative Interaction*. Lausanne: Frontiers Media SA.

Holler, J., H. Shovelton, and G. Beattie
2009. Do Iconic Hand Gestures Really Contribute to the Communication of Semantic Information in a Face-to-Face Context? *Journal of Nonverbal Behavior*, 33(2):73–88.

Hostetter, A. B. and M. W. Alibali
2008. Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15(3):495–514.

Hothorn, T., K. Hornik, and A. Zeileis
2006. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15(3):651–674.

Hsieh, C.-Y. C. and L. I-Wen Su
2019. Construction in conversation: An Interactional Construction Grammar approach to the use of xiangshuo 'think' in spoken Taiwan Mandarin. *Review of Cognitive Linguistics*, 17(1):131–154.

Huang, Y. and F. Ferreira
2020. The Application of Signal Detection Theory to Acceptability Judgments. *Frontiers in Psychology*, 11:73.

Hutchins, E.
1995. *Cognition in the wild*. Cambridge, MA: MIT Press.

Hutchins, E.
2006. The distributed cognition perspective on human interaction. In *Roots of Human Sociality: Culture, Cognition and Interaction*, S. C. Levinson and N. J. Enfield, eds. London: Berg.

Hymes, D.
1967. Models of the Interaction of Language and Social Setting. *Journal of Social Issues*, 23(2):8–28.

Im, S. and S. Baumann
2020. Probabilistic relation between co-speech gestures, pitch accents and information status. *Proceedings of the Linguistic Society of America*, 5(1):685.

Imo, W.
2015. Interactional Construction Grammar. *Linguistics Vanguard*, 1(1):69–77.

Janda, L. A.
2003. Russian Aspect at Your Fingertips: Why What You Know about Matter Matters, or a User-Friendly Conceptualization of Verbal Aspect in Russian. *The Slavic and East European Journal*, 47(2):251–281.

Janda, L. A., ed.
2013a. *Cognitive linguistics: the quantitative turn: the essential reader*. Berlin: De Gruyter Mouton.

Janda, L. A., ed.
2013b. *Why Russian aspectual prefixes aren't empty: prefixes as verb classifiers*. Bloomington, IN: Slavica. OCLC: 829974508.

Janda, L. A.
2015. Russian Aspectual Types: Croft's Typology Revised. In *Studies in Slavic Linguistics and Accentology in Honor of Ronald F. Feldstein*, M. Shrager, G. Fowler, S. Franks, and E. Andrews, eds., Pp. 146–167. Bloomington, IN: Slavica Publishers.

Jannedy, S. and N. Mendoza-Denton
2005. Structuring Information through Gesture and Intonation. *Interdisciplinary Studies on Information Structure*, 3:199–244.

Jantunen, T.
2015. How long is the sign? *Linguistics*, 53(1):93–124.

Januška, J.
2017. *Porovnávání středoevropských jazyků: za horizont strukturních rysů a lexikálních přejímek*. Doctoral dissertation, Charles University, Prague.

Jefferson, G.
2004. Glossary of transcript symbols with an introduction. In *Conversation analysis: studies from the first generation*, G. H. Lerner, ed., Pp. 13–31. Amsterdam: John Benjamins.

Jehlička, J.
2016. Alignment of sentence focus and gesture in spontaneous English conversations. Presented at 7th Conference of the International Society for Gesture Studies, Paris.

Jehlička, J. and E. Lehečková
2020. Multimodal Event Construals: The Role of Co-Speech Gestures in English vs. Czech Interactions. *Zeitschrift für Anglistik und Amerikanistik*, 68(4):351–377.

Johnson, M.
1987. *The body in the mind: the bodily basis of meaning, imagination, and reason*. Chicago, IL: University of Chicago Press. OCLC: 879576101.

Juhasz, B. J. and M. J. Yap
2013. Sensory experience ratings for over 5,000 mono- and disyllabic words. *Behavior Research Methods*, 45(1):160–168.

Kaderka, P. and Z. Svobodová
2006. Jak přepisovat audiovizuální záznam rozhovoru? Manuál pro přepisovatele televizních diskusních pořadů. *Jazykovědné aktuality*, 43(3-4):18–51.

Karpiński, M., E. Jarmolowicz, Z. Malisz, M. Szczyszek, and K. Juszczyk
2008. Rejestracja, transkrypcja i tagowanie mowy oraz gestów w narracji dzieci i dorosłych. In *Investigationes Linguisticae, vol. XVI*, Pp. 83–89.

Kelly, S., A. Bailey, and Y. Hirata
2017. Metaphoric Gestures Facilitate Perception of Intonation More than Length in Auditory Judgments of Non-Native Phonemic Contrasts. *Collabra: Psychology*, 3(1):7.

Kelly, S. D.
2017. Exploring the boundaries of gesture-speech integration during language comprehension. In *Gesture Studies*, R. B. Church, M. W. Alibali, and S. D. Kelly, eds., volume 7, Pp. 243–265. Amsterdam: John Benjamins.

Kelly, S. D., A. Özyürek, and E. Maris
2010. Two Sides of the Same Coin: Speech and Gesture Mutually Interact to Enhance Comprehension. *Psychological Science*, 21(2):260–267.

Kemmerer, D.
2015. Does the motor system contribute to the perception and understanding of actions? Reflections on Gregory Hickok's The myth of mirror neurons: the real neuroscience of communication and cognition. *Language and Cognition*, 7(03):450–475.

Kemmerer, D., B. Chandrasekaran, and D. Tranel
2007. A case of impaired verbalization but preserved gesticulation of motion events. *Cognitive Neuropsychology*, 24(1):70–114.

Kendon, A.
1972. Some relationships between body motion and speech. An analysis of an example. In *Studies in Dyadic Communication*, A. W. Siegman and B. Pope, eds., Pp. 177–210. Elmsford, NY: Pergamon Press.

Kendon, A.
1980. Gesticulation and speech: two aspects of the process of utterance. In *The Relationship of Verbal and Nonverbal Communication*, M. R. Key, ed. The Hague: Mouton.

Kendon, A.
1988. How gestures can become like words. In *Cross-Cultural Perspectives in Nonverbal Communication*, F. Poyatos, ed., Pp. 131–141. Toronto: Hogrefe.

Kendon, A.
1994. Do Gestures Communicate? A Review. *Research on Language and Social Interaction*, 27:175–200.

Kendon, A.
2004. *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.

Kendon, A.
2013. Exploring the utterance roles of visible bodily action: A personal account. In *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. (Handbooks of Linguistics and Communication Science 38.1.)*, Pp. 7–28. Berlin: De Gruyter Mouton.

Kendrick, K. H. and F. Torreira
2015. The Timing and Construction of Preference: A Quantitative Study. *Discourse Processes*, 52(4):255–289.

Kita, S.
2000. How representational gestures help speaking. In *Language and Gesture*, D. McNeill, ed., Pp. 162–185. Cambridge: Cambridge University Press.

Kita, S.
2003. Interplay of gaze, hand, torso orientation, and language in pointing. In *Poiting: Where Language, Culture, and Cognition Meet*, S. Kita, ed. Mahwah, NJ: Lawrence Erlbaum Associates.

Kita, S., I. van Gijn, and H. van der Hulst
1998. Movement phases in signs and co-speech gestures, and their transcription by human coders. In *Gesture and Sign Language in Human-Computer Interaction*, J. G. Carbonell, J. Siekmann, G. Goos, J. Hartmanis, J. van Leeuwen, I. Wachsmuth, and M. Fröhlich, eds., volume 1371, Pp. 23–35. Berlin: Springer.

Kita, S. and A. Özyürek
2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48(1):16–32.

Kita, S., A. Özyürek, S. Allen, A. Brown, R. Furman, and T. Ishizuka
2007. Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes*, 22(8):1212–1236.

Klima, E. S. and U. Bellugi
1979. *The signs of language*. Cambridge, MA: Harvard Univ. Press. OCLC: 837627301.

Knop, S. D.
2020. Expressions of motion events in German: an integrative constructionist approach for FLT1. *CogniTextes*, (21).

Koffka, K.
1935. *Principles of Gestalt psychology*. New York, NY: Harcourt - Brace.

Kok, K. I.
2016. *The status of gesture in cognitive-functional models of grammar*. Doctoral dissertation, Vrije Universiteit Amsterdam, Amsterdam.

Kokorniak, I.
2017. Cross-linguistic aspectual variation and the mental predicate think: The case of English and Polish. Presented at ICLC 14, Tartu.

Kopečný, F.
1962. *Slovesný vid v češtině*. Praha: ČSAV.

Kragh, H.
2002. *Quantum generations: a history of physics in the twentieth century*, 5th printing, and 1. paperback printing edition. Princeton, N.J.: Princeton University Press. OCLC: 248763258.

Krahmer, E. and M. Swerts
2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57(3):396–414.

Krauss, R. M., Y. Chen, and R. F. Gottesman
2000. Lexical gestures and lexical access: a process model. In *Language and Gesture*, Pp. 261–283. New York, NY: Cambridge University Press.

Krifka, M.
1992. Thematic relations as links between nominal reference and temporal constitution. In *Lexical Matters*, I. A. Sag and A. Szabolcsi, eds., Pp. 29–53. Stanford, CA: CSLI.

Kuhn, J.
2017. Telicity and iconic scales in ASL [manuscript].

Kurby, C. A. and J. M. Zacks
2008. Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2):72–79.

Křen, M., V. Cvrček, T. Čapka, A. Čermáková, and M. Hnátková
2016. SYN2015: Representative Corpus of Contemporary Written Czech. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16),*, Pp. 2522–2528, Portorož. ELRA.

Kříž, A. and F. Smolík
2015. Hodnocení představitelnosti, konkrétnosti, specifičnosti, familiarity a věku osvojení českých substantiv a sloves: vztahy a souvislosti. *Československá psychologie*, 59(6):507–520.

Laban, R. v. and F. C. Lawrence
1947. *Effort*. London: Macdonald & Evans. OCLC: 316199.

Labov, W.
1964. Phonological Correlates of Social Stratification. *American Anthropologist*, 66(6):164–176.

Laing, C.
2019. A role for onomatopoeia in early language: evidence from phonological development. *Language and Cognition*, 11(02):173–187.

Lakoff, G.
1977. Linguistic gestalts. *Chicago Linguistic Society*, 13:236–287.

Lakoff, G.
1987. *Women, fire, and dangerous things: what categories reveal about the mind*, paperback ed., [nachdr.] edition. Chicago, IL: The Univ. of Chicago Press. OCLC: 935630820.

Lakoff, G. and M. Johnson
1980. *Metaphors we live by*. Chicago, IL: University of Chicago Press. OCLC: 899007727.

Lambrecht, K.
1994. *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*, transf. to digital print edition. Cambridge: Cambridge Univ. Press.

Landis, R. J. and G. K. Koch
1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174.

Lane, H. L.
1992. *The mask of benevolence: disabling the deaf community*. New York, NY: Knopf.

Langacker, R. W.
1987a. *Foundations of Cognitive Grammar I: Theoretical Prerequisities*. Stanford, CA: Stanford University Press.

Langacker, R. W.
1987b. Nouns and verbs. *Language*, 63(1):53–94.

Langacker, R. W.
2017. *Ten lectures on the elaboration of cognitive grammar*, number 18. Leiden: Brill.

Leech, G. N.
1983. *Principles of Pragmatics*. New York, NY: Longman. OCLC: 9043896.

Lehečková, E.
2011. *Teličnost a skalárnost deadjektivních sloves v češtině*. Doctoral dissertation, Charles University, Prague.

Lehečková, E.
under review. Teličnost v kognitivní lingvistice a její uplatnění v empirickém výzkumu (Telicity in Cognitive Linguistics and its Application in Empirical Research).

Lehečková, E. and J. Jehlička
2018. Multimodální konstrukce: jazyk a gestikulace jako vtělená kognice. *Studie z aplikované lingvistiky / Studies in applied linguistics*, 9(2):89–103.

Lehečková, E. and J. Jehlička
2019. Gestikulace ve sdíleném prostoru jako kooperativní utváření významu. *Časopis pro moderní filologii*, 101(2):150–169.

Lehečková, E., J. Jehlička, and M. Zíková
2019. Interplay of information structure, pitch contour, and gesture in spontaneous interactions. Presented at ICLC 15, Nishinomiya.

Lemaitre, G., H. Scurto, J. Françoise, F. Bevilacqua, O. Houix, and P. Susini
2017. Rising tones and rustling noises: Metaphors in gestural depictions of sounds. *PLOS ONE*, 12(7):e0181786.

Lenneberg, E. H.
1967. *Biological foundations of language*. New York, NY: Wiley. OCLC: 557223.

Leonard, T. and F. Cummins
2011. The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26(10):1457–1471.

Levelt, W. J. M.
1989. *Speaking: From Intention to Articulation.* Cambridge, MA: The MIT Press.

Levelt, W. J. M.
2013. *A history of psycholinguistics: the pre-Chomskyan era*. Oxford: Oxford University Press.

Levin, B.
1993. *English verb classes and alternations: a preliminary investigation*. Chicago, IL: University of Chicago Press.

Levinson, S. C.
1996a. Frames of reference and Molyneux's question: Cross-linguistic evidence — Max Planck Institute for Psycholinguistics. In *Language and space*, P. Bloom, M. Peterson, and L. Nadel, eds., Pp. 109–169. Cambridge, MA: MIT Press.

Levinson, S. C.
1996b. Relativity in spatial conception and description. In *Rethinking Linguistic Relativity*, J. J. Gumperz and S. C. Levinson, eds., Pp. 177–202. Cambridge: Cambridge University Press.

Levinson, S. C.
2003. *Space in language and cognition: explorations in cognitive diversity*, number 5. Cambridge: Cambridge University Press.

Levinson, S. C. and J. Holler
2014. The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651):20130302.

Levshina, N.
2015. *How to do linguistics with R: data exploration and statistical analysis*. Amsterdam: John Benjamins.

Leys, C., C. Ley, O. Klein, P. Bernard, and L. Licata
2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766.

Liddell, S.
2000. Indicating verbs and pronouns: Pointing away from agreement. In *The Signs of Language revisited: An anthology to honor Ursula Bellugi and Edward Klima*, K. Emmorey and H. Lane, eds., Pp. 303–320. Mahwah, NJ: Lawrence Erlbaum Associates.

Linell, P.
1982. *The written language bias in linguistics*, number 2) in (SIC [Studies in communication]. Linköping: University of Linköping. OCLC: 251790081.

Linell, P.
2005. *The written language bias in linguistics: its nature, origins, and transformations*. London ; New York: Routledge.

Linell, P.
2009. Grammatical constructions in dialogue. In *Constructional Approaches to Language*, A. Bergs and G. Diewald, eds., Pp. 97–110. Amsterdam: John Benjamins.

Lis, M.
2014. *Multimodal representation of entities: A corpus-based investigation of co-speech hand gesture*. Doctoral dissertation, University of Copenhagen, Copenhagen.

Lis, M. and C. Navarretta
2013. Classifying the form of iconic hand gestures from the linguistic categorization of co-occurring verbs. In *Proceedings from the 1st European Symposium on Multimodal Communication, University of Malta, Valletta, October 17–18, 2013*, Pp. 41–50.

Liszkowski, U., P. Brown, T. Callaghan, A. Takada, and C. de Vos
2012. A Prelinguistic Gestural Universal of Human Communication. *Cognitive Science*, 36(4):698–713.

Loehr, D. P.
2004. *Gesture and intonation*. Doctoral dissertation, Goergetown University, Washington, D.C.

Loehr, D. P.
2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, 3(1):71–89.

Lotman, J. M.
1984. O semiosfere. *Trudy po znakovym sistemam*, 17:5–23.

Lucy, J. A.
1992. *Language diversity and thought: a reformulation of the linguistic relativity hypothesis*. Cambridge; New York, NY: Cambridge University Press.

Lücking, A., K. Bergmann, F. Hahn, S. Kopp, and H. Rieser
2010. The Bielefeld Speech and Gesture Alignment Corpus (SaGA). *Proceedings of the LREC 2010 Workshop "Multimodal Corpora – Advances in Capturing, Coding and Analyzing Multimodality"*, Pp. 92–98.

Macaulay, R. K. S.
1978. Review of Comrie (1976) and Friedrich (1974). *Language*, 54(2):416–420.

MacWhinney, B., D. Fromm, M. Forbes, and A. Holland
2011. AphasiaBank: Methods for studying discourse. *Aphasiology*, 25(11):1286–1307.

Makowski, D.
2018. The psycho Package: an Efficient and Publishing-Oriented Workflow for Psychological Science. *The Journal of Open Source Software*, 3(22):470.

Malaia, E. and R. B. Wilbur
2012. Kinematic Signatures of Telic and Atelic Events in ASL Predicates. *Language and Speech*, 55(3):407–421.

Mandel, M.
1977. Iconic devices in American Sign Language. In *On the Other Hand: New Perspectives on American Sign Language.*, L. A. Friedman, ed., Pp. 57–108. New York, NY: Academic Press.

Marstaller, L. and H. Burianová
2014. The multisensory perception of co-speech gestures – A review and meta-analysis of neuroimaging studies. *Journal of Neurolinguistics*, 30:69–77.

Martinet, A.
1960. *Éléments de linguistique générale*. Paris: Armand Colin.

Masson-Carro, I., M. Goudbeek, and E. Krahmer
2017. How What We See and What We Know Influence Iconic Gesture Production. *Journal of Nonverbal Behavior*, 41(4):367–394.

Matlock, T.
2004. Fictive motion as cognitive simulation. *Memory & Cognition*, 32(8):1389–1400.

McClave, E.
1998. Pitch and Manual Gestures. *Journal of Psycholinguistic Research*, 27(1):69–89.

McGurk, H. and J. MacDonald
1976. Hearing lips and seeing voices. *Nature*, 264(5588):746–748.

McNeill, D.
1979. *The conceptual basis of language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

McNeill, D.
1992. *Hand and mind: what gestures reveal about thought*. Chicago, IL: University of Chicago Press.

McNeill, D.
2005. *Gesture and Thought*. Chicago, IL: University of Chicago Press.

McNeill, D.
2006. Gesture and Communication. In *Encyclopedia of Language & Linguistics*, K. Brown, ed., Pp. 58–66. Amsterdam: Elsevier.

McNeill, D.
2012. *How Language Began*. Cambridge: Cambridge University Press.

McNeill, D.
2013. The growth point hypothesis of language and gesture as a dynamic and integrated system. In *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. (Handbooks of Linguistics and Communication Science 38.1.)*, Pp. 135–155. Berlin: De Gruyter Mouton.

McNeill, D.
2017. Gesture-speech unity: What it is, where it came from. In *Why Gesture? How the Hands Fuction in Speaking, Thinking and Communicating*, R. B. Church, M. W. Alibali, and S. D. Kelly, eds., Pp. 77–101. Amsterdam: John Benjamins.

McNeill, D., J. Cassell, and K.-E. McCullough
1994. Communicative Effects of Speech-Mismatched Gestures. *Research on Language & Social Interaction*, 27(3):223–237.

McNeill, D. and E. T. Levy
1982. Conceptual Representations in Language Activity and Gesture. In *Speech, Place, and Action. Studies in Deixis and Related Topics*, R. J. Jarvella and W. Klein, eds., Pp. 271–295. Chichester et al.: John Wiley & Sons.

McNeill, D., E. T. Levy, and S. D. Duncan
2015. Gesture in Discourse. In *The Handbook of Discourse Analysis*, D. Tannen, H. E. Hamilton, and D. Schiffrin, eds., Pp. 262–289. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Meir, I.
2001. Verb classifiers as noun incorporation in Israeli sign language. In *Yearbook of morphology 1999*, G. E. Booij and J. van Marle, eds., Pp. 299–319. Dordrecht: Kluwer Academic Publishers.

Merleau-Ponty, M.
1945. *Phénoménologie de la perception*. Paris: Gallimard.

Merleau-Ponty, M.
1962. *Phenomenology of perception*. London : Routledge & K. Paul ; New York : Humanities Press.

Mertins, B.
2018. *Sprache und Kognition: Ereigniskonzeptualisierung im Deutschen und Tschechischen*. Berlin, Boston: De Gruyter.

Miranda, M. A. and P. H. A. Mendes
2015. The role of gestures in the construction of multimodal metaphors: analysis of a political-electoral debate. *Revista Brasileira de Linguística Aplicada*, 15(2):343–376.

Mittelberg, I.
2008. Peircean semiotics meets conceptual metaphor. In *Metaphor and Gesture*, A. J. Cienki and C. Müller, eds., Pp. 115–154. Amsterdam: John Benjamins.

Mittelberg, I.
2013. The exbodied mind: Cognitive-semiotic principles as motivating forces in gesture. In *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction. Handbooks of Linguistics and Communication Science (38.1)*, C. Müller, A. J. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and S. Tessendorf, eds., Pp. 750–779. Berlin: Mouton de Gruyter.

Mittelberg, I.
2014. Gestures and iconicity. In *Body - Language - Communication (HSK 38.2)*, C. Müller, A. J. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem, eds., Pp. 1712–1732. Berlin: De Gruyter Mouton.

Mittelberg, I.
2018. Gestures as image schemas and force gestalts: A dynamic systems approach augmented with motion-capture data analyses. *Cognitive Semiotics*, 11(1):20180002.

Mittelberg, I. and V. Evola
2014. Iconic and representational gestures. In *Body - Language - Communication (HSK 38.2)*, C. Müller, A. J. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem, eds., Pp. 1732–1746. Berlin: De Gruyter.

Mittelberg, I. and L. R. Waugh
2009. Metonymy first, metaphor second: A cognitivesemiotic approach to multimodal figures of thought in co-speech gesture. In *Multimodal Metaphor*, C. J. Forceville and E. Urios-Aparisi, eds., Pp. 329–356. Berlin: Mouton de Gruyter.

Monaghan, P., R. C. Shillcock, M. H. Christiansen, and S. Kirby
2014. How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651):20130299.

Morett, L. M. and L.-Y. Chang
2015. Emphasising sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, 30(3):347–353.

Müller, C.
1998. Iconicity and gesture. In *Oralité et gestualité*, S. Santi, ed., Pp. 321–328. Montréal: L'Harmattan.

Newport, E. and T. Supalla
1980. The structuring of language: Clues from the acquisition of signed and spoken language. In *Signed and spoken language: Biological constraints on linguistic form*, U. Bellugi and M. Studdert-Kennedy, eds., Pp. 187–211. Weinheim: Verlag Chemie.

Ningelgen, J. and P. Auer
2017. Is there a multimodal construction based on non-deictic so in German? *Linguistics Vanguard*, 3(s1).

Nir, B., G. Dori-Hacohen, and Y. Maschler
2014. Formulations on Israeli political talk radio: From actions and sequences to stance via dialogic resonance. *Discourse Studies*, 16(4):534–571.

Occhino, C., B. Anible, E. Wilkinson, and J. P. Morford
2017. Iconicity is in the eye of the beholder: How language experience affects perceived iconicity. *Gesture*, 16(1):100–126.

O'Connor, R.
2008. A prosodic projection for Role and Reference Grammar. In *Investigations of the Syntax–Semantics–Pragmatics Interface*, Pp. 228–284. Amsterdam: John Benjamins.

Ortega, G. and A. Özyürek
2019. Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. *Behavior Research Methods*, 52:51–67.

Özyürek, A., S. Kita, S. Allen, R. Furman, and A. Brown
2005. How does linguistic framing of events influence co-speech gestures?: Insights from crosslinguistic variations and similarities. *Gesture*, 5(1-2):219–240.

Özyürek, A., R. M. Willems, S. Kita, and P. Hagoort
2007. On-line Integration of Semantic Information from Speech and Gesture: Insights from Event-related Brain Potentials. *Journal of Cognitive Neuroscience*, 19(4):605–616.

Pagán Cánovas, C., J. Valenzuela, D. Alcaraz Carrión, I. Olza, and M. Ramscar
2020. Quantifying the speech-gesture relation with massive multimodal datasets: Informativity in time expressions. *PLOS ONE*, 15(6):e0233892.

Paivio, A., J. Yuille, and S. Madigan
1968. Concreteness, imagery, and meaningfulness norms for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, (76):1–25.

Papafragou, A., J. Hulbert, and J. Trueswell
2008. Does language guide event perception? Evidence from eye movements. *Cognition*, 108(1):155–184.

Parrill, F.
2010. Viewpoint in speech–gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes*, 25(5):650–668.

Parrill, F., B. K. Bergen, and P. V. Lichtenstein
2013. Grammatical aspect, gesture, and conceptualization: Using co-speech gesture to reveal event representations. *Cognitive Linguistics*, 24(1):135–158.

Parsons, T.
1980. Modifiers and Quantifiers in Natural Language. *Canadian Journal of Philosophy*, 6:29–60.

Pashler, H. and E. Wagenmakers
2012. Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*, 7(6):528–530.

Peirce, C. S.
1931. *Collected Papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Peirce, C. S.
1932. *Collected Writings 2: Elements of Logic*. Cambridge, MA: Harvard University Press.

Peirce, J. W.
2007. PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2):8–13.

Perlman, M.
2017. Debunking two myths against vocal origins of language: Language is iconic and multimodal to the core. *Interaction Studies*, 18(3):376–401.

Perniss, P.
2007. *Space and Iconicity in German Sign Language (DGS).* Doctoral dissertaion, Radboud Universiteit, Nijmegen.

Perniss, P., J. C. Lu, G. Morgan, and G. Vigliocco
2018. Mapping language to the world: the role of iconicity in the sign language input. *Developmental Science*, 21(2):e12551.

Perniss, P., R. L. Thompson, and G. Vigliocco
2010. Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology*, 1:227.

Perniss, P. and G. Vigliocco
2014. The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651):20130300.

Perry, L. K., M. Perlman, and G. Lupyan
2015. Iconicity in English and Spanish and Its Relation to Lexical Category and Age of Acquisition. *PLoS ONE*, 10(9):1–17.

Pike, K. L.
1967. *Language in relation to a unified theory of the structure of human behavior*. The Hague: Mouton. OCLC: 885116357.

Pinker, S.
1994. *The language instinct*. New York, NY: HarperCollins Publishers. OCLC: 894474121.

Popper, K.
1935. *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft*. Vienna: Springer Verlag.

Pouw, W., S. J. Harrison, and J. A. Dixon
2020a. Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony. *Journal of Experimental Psychology: General*, 149(2):391–404.

Pouw, W., S. J. Harrison, N. Esteve-Gibert, and J. A. Dixon
2020b. Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures. *The Journal of the Acoustical Society of America*, 148(3):1231–1247.

Prieto, P., A. Cravotta, O. Kushch, P. Rohrer, and I. Vilà-Giménez
2018. Deconstructing beat gestures: a labelling proposal. In *9th International Conference on Speech Prosody 2018*, Pp. 201–205. ISCA.

R Core Team
2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Radvansky, G. A. and J. M. Zacks
2014. *Event cognition*. Oxford ; New York: Oxford University Press.

Rizzolatti, G. and M. A. Arbib
1998. Language within our grasp. *Trends in Neurosciences*, 21(5):188–194.

Ruby, J.
1980. Franz Boas and early camera study of behavior. *The Kinesis Report*, 3(6-11):16.

Ruth-Hirrel, L. and S. Wilcox
2018. Speech-gesture constructions in cognitive grammar: The case of beats and points. *Cognitive Linguistics*, 29(3):453–493.

Sacks, H., E. A. Schegloff, and G. Jefferson
1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4):696–735.

Sasse, H.-J.
2002. Recent activity in the theory of aspect: Accomplishments, achievements, or just non-progressive state? *Linguistic Typology*, 6(2):199–271.

Schalber, K.
2006. Event visibility in Austrian Sign Language (ÖGS). *Sign Language & Linguistics*, 9(1/2):207–231.

Schegloff, E. A.
1968. Sequencing in Conversational Openings. *American Anthropologist*, 70(6):1075–1095.

Schegloff, E. A.
1985. On some gestures' relation to talk. In *Structures of Social Action*, J. M. Atkinson, ed., Pp. 266–296. Cambridge: Cambridge University Press.

Schembri, A., K. Cormier, and J. Fenlon
2018. Indicating verbs as typologically unique constructions: Reconsidering verb 'agreement' in sign languages. *Glossa: a journal of general linguistics*, 3(1):89.

Schmid, H.-J.
2016. Why Cognitive Linguistics must embrace the social and pragmatic dimensions of language and how it could do so more seriously. *Cognitive Linguistics*, 0(0).

Schoonjans, S.
2014. *Modalpartikeln als multimodale Konstruktionen. Eine korpusbasierte Kookkurrenzanalyse von Modalpartikeln und Gestik im Deutschen*. Doctoral dissertation, KU Leuven, Leuven.

Schoonjans, S.
2017. Multimodal Construction Grammar issues are Construction Grammar issues. *Linguistics Vanguard*, 3(s1).

Searle, J. R.
1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, MA: Cambridge University Press.

Sekine, K., M. L. Rose, A. M. Foster, M. C. Attard, and L. E. Lanyon
2013. Gesture production patterns in aphasic discourse: In-depth description and preliminary predictions. *Aphasiology*, 27(9):1031–1049.

Silverman, K., M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg
1992. TOBI: a standard for labeling English prosody. *Proceedings of ICSLP 1992,*, Pp. 867–870.

Sinha, C.
2007. Cognitive Linguistics, Psychology and Cognitive Science. In *The Oxford Handbook of Cognitive Linguistics*, D. Geeraerts and H. Cuyckens, eds., Pp. 1266–1294. Oxford: Oxford University Press.

Skinner, B. F.
1957. *Verbal behavior*. New York, NY: Appleton-Century-Crofts, Inc. OCLC: 932081700.

Slobin, D. I.
1996. From "thought to language" to "thinking for speaking". In *Rethinking Linguistic Relativity*, J. J. Gumperz and S. C. Levinson, eds., Pp. 70–96. Cambridge: Cambridge University Press.

Slobin, D. I.
2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In *Relating Events in Narrative: Vol: 2. Typological and Contextual Perspectives*, S. Strömqvist and L. Verhoeven, eds., Pp. 219–257. Mahwah, NJ: Lawrence Erlbaum Associates.

Smith, C. S.
1997. *The parameter of aspect*, number 43, 2. ed edition. Dordrecht: Kluwer Academic Publ. OCLC: 833221703.

So, W. C., A. Low, D. F. Yap, E. Kheng, and M. Yap
2013. Iconic gestures prime words: comparison of priming effects when gestures are presented alone and when they are accompanying speech. *Frontiers in Psychology*, 4:779.

Spreen, O. and R. W. Schulz
1966. Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. *Journal of Verbal Learning and Verbal Behavior*, 5(5):459–468.

Starý Kořánová, I.
2019. *Vidová kolokabilita*. Doctoral dissertation, Charles University, Prague.

Steels, L.
2017. Basics of Fluid Construction Grammar. *Constructions and Frames*, 9(2):178–225.

Stefanowitsch, A. and S. T. Gries
2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2):209–243.

Stokoe, W.
1960. *Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf*. Buffalo, NY: University of Buffalo.

Streeck, J.
2009. *Gesturecraft: the manu-facture of meaning*, number v. 2. Amsterdam: John Benjamins. OCLC: ocn276333954.

Streeck, J.
2015. Embodiment in Human Communication. *Annual Review of Anthropology*, 44(1):419–438.

Strickland, B., C. Geraci, E. Chemla, P. Schlenker, M. Kelepir, and R. Pfau
2015. Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences*, 112(19):5968–5973.

Strobl, C., A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis
2008. Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1):307.

Sweetser, E.
2012. Introduction: viewpoint and perspective in language and gesture, from the Ground down. In *Viewpoint in Language: A Multimodal Perspective*, B. Dancygier and E. Sweetser, eds., Pp. 1–22. Cambridge: Cambridge University Press.

Swets, J. A., W. P. Tanner, Theodore, G. Birdsall, H. R. Blackwell, and W. M. K. For
1961. Decision processes in perception. *Psychological Review*, 68:301–340.

Tagliamonte, S. A. and R. H. Baayen
2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change*, 24(2):135–178.

Talmy, L.
1972. *Semantic Structures in English and Atsugewi*. Doctoral dissertation, University of California, Berkeley, CA.

Talmy, L.
1985. Lexicalization patterns: semantic structure in lexical forms. In *Language typology and syntactic description, Vol.3, Grammatical categories and the lexicon*, T. Shopen, ed., Pp. 57–149. Cambridge: Cambridge University Press.

Talmy, L.
2000. *Toward a cognitive semantics*. Cambridge, MA: MIT Press.

Talmy, L.
2017. *The Targeting System of Language*. Cambridge, MA: The MIT Press.

Thierry, G., P. Athanasopoulos, A. Wiggett, B. Dering, and J.-R. Kuipers
2009. Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, 106(11):4567–4570.

Tomasello, M.
2005. *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press. OCLC: 254708552.

Traugott, E. C. and G. Trousdale
2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Trueswell, J. C. and A. Papafragou
2010. Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, 63(1):64–82.

Turk, O.
2020. *Gesture, Prosody and Information Structure Synchronisation in Turkish*. Doctoral dissertation, Victoria University of Wellington, Wellington.

Turner, M. B. and F. F. Steen
2013. Multimodal Construction Grammar. In *Language and the Creative Mind*, M. Borkent, B. Dancygier, and J. Hinnel, eds., Pp. 255–274. Stanford, CA; Chicago, IL: CSLI Publications & University of Chicago Press.

Tversky, B. and A. Jamalian
2012. Gestures Alter Thinking About Time. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society (CogSci 2012)*, N. Miyake, D. Peebles, and R. P. Cooper, eds., Pp. 551–557, Austin, TX. Cognitive Science Society.

Uexküll, J. v.
1992. A stroll through the worlds of animals and men: A picture book of invisible worlds. *Semiotica*, 89(4):319–391.

Uhrig, P.
2019. Theoretical and practical aspects of crossmodal collostructions. Presented at Gesture-Sign Workshop Prague 2019, Prague.

Van Valin, R. D.
2005. *Exploring the syntax-semantics interface*. Cambridge; New York: Cambridge University Press. OCLC: 69953250.

VanTrijp, R.
2013. A comparison between Fluid Construction Grammar and Sign-Based Construction Grammar. *Constructions and Frames*, 5(1):88–116.

Vendler, Z.
1967. *Linguistics in philosophy*. Ithaca, NY: Cornell University Press. OCLC: 258761031.

Verhagen, A.
2005. *Constructions of intersubjectivity: discourse, syntax, and cognition*. New York, NY: Oxford University Press.

Verhagen, A.
2015. Grammar and cooperative communication. In *Handbook of Cognitive Linguistics*, E. Dabrowska and D. Divjak, eds. Berlin: De Gruyter.

Veselý, L.
2014. *Gramatické studie 1: Příspěvky k české aspektologii*. Olomouc: Univerzita Palackého v Olomouci. OCLC: 907520526.

Vigliocco, G., Y. Zhang, N. Del Maschio, R. Todd, and J. Tuomainen
2020. Electrophysiological signatures of English onomatopoeia. *Language and Cognition*, 12(1):15–35.

Vinson, D. P., K. Cormier, T. Denmark, A. Schembri, and G. Vigliocco
2008. The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods*, 40(4):1079–1087.

Volterra, V., O. Capirci, M. C. Caselli, P. Rinaldi, and L. Sparaci
2017. Developmental evidence for continuity from action to gesture to sign/word. *Language, Interaction and Acquisition*, 8(1):13–41.

von Stutterheim, C., M. Andermann, M. Carroll, M. Flecken, and B. Schmiedtová
2012. How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics*, 50(4):33–867.

Vygotsky, L. S.
1986. *Thought and language*. Cambridge, MA: MIT Press.

Warner-Garcia, S.
2013. Gestural Resonance: The Negotiation of Differential Form and Function in Embodied Action. *Crossroads of Language, Interaction and Culture*, 9(1):55–78.

Wessel-Tolvig, B. and P. Paggio
2016. Revisiting the thinking-for-speaking hypothesis: Speech and gesture representation of motion in Danish and Italian. *Journal of Pragmatics*, 99:39–61.

Westermann, D. H.
1937. Laut Und Sinn in Einigen westafrikanischen Sprachen. *Archiv Für Vergleichende Phonetik*, 1:154–72, 193–211.

Whorf, B. L.
1941. The relation of habitual thought and behavior to language. In *Language, Culture, and Personality: Essays in memory of Edward Sapir*, L. Spier, I. A. Hallowell, and S. S. Newman, eds., Pp. 75–93. Menasha, WI: Sapir Memorial Publication Fund.

Whorf, B. L.
1956. *Language, Thought, and Reality*. Cambridge, MA: Technology Press of Massachusetts Institute of Technology.

Wide, C.
2009. Interactional Construction Grammar: Contextual features of determination in dialectal Swedish. In *Contexts and Constructions*, A. Bergs and G. Diewald, eds., Pp. 111–142. Amsterdam: John Benjamins.

Wilbur, R.
2003. Representations of telicity in ASL. *Chicago Linguistic Society*, 39:354–368.

Wilbur, R.
2008. Complex Predicates Involving Events, Time and Aspect: Is This Why Sign Languages Look so Similar? In *Signs of the Time: Selected Papers from TISLR 2004.*, J. Quer, ed., Pp. 217–250. Hamburg: Signum.

Wilcox, S.
2004. Cognitive iconicity: conceptual spaces, meaning, and gesture in signed languages. *Cognitive Linguistics*, 15(2):119–147.

Wilcox, S.
2018. Cognitive Iconicity, Conceptual Spaces, Meaning, and Gesture. In *Ten Lectures on Cognitive Linguistics and the Unification of Spoken and Signed Languages*, Pp. 67–87. Leiden: Brill.

Wilcox, S. and J. P. Morford
2007. Empirical methods in signed language research. In *Methods in Cognitive Linguistics*, M. Gonzalez Marquez, I. Mittelberg, S. Coulson, and M. J. Spivey, eds., Pp. 173–202. Amsterdam: John Benjamins.

Wilkins, D.
2003. Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In *Pointing: Where Language, Culture, and Cognition Meet*, S. Kita, ed., Pp. 171–215. Mahwah, NJ: Lawrence Erlbaum Associates.

Willems, R. M., L. Labruna, M. D'Esposito, R. Ivry, and D. Casasanto
2011. A Functional Role for the Motor System in Language Understanding: Evidence From Theta-Burst Transcranial Magnetic Stimulation. *Psychological Science*, 22(7):849–854.

Willems, R. M., A. Özyürek, and P. Hagoort
2007. When Language Meets Action: The Neural Integration of Gesture and Speech. *Cerebral Cortex*, 17(10):2322–2333.

Wilson, M.
2002. Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–636.

Winter, B.
2019. *Statistics for Linguists: An Introduction Using R*, 1 edition. New York, NY: Routledge.

Winter, B., M. Perlman, L. K. Perry, and G. Lupyan
2017. Which words are most iconic? Iconicity in English sensory words. *Interaction Studies*, 18(3):430–451.

Wittenburg, P., H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes
2006. Elan: a professional framework for multimodality research. In *In Proceedings of Language Resources and Evaluation Conference (LREC*.

Wundt, W.
1900a. *Die Sprache*. Leipzig: Wilhelm Engelmann.

Wundt, W.
1900b. *Völkerpsychologie: Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythus und Sitte*. Leipzig: Wilhelm Engelmann.

Wundt, W.
1973. *Language of Gestures*. The Hague: Mouton.

Yang, C. L., H. Zhang, H. Duan, and H. Pan
2019. Linguistic Focus Promotes the Ease of Discourse Integration Processes in Reading Comprehension: Evidence From Event-Related Potentials. *Frontiers in Psychology*, 9:2718. Publisher: Frontiers.

Yoshioka, K.
2008. Gesture and information structure in first and second language. *Gesture*, 8(2):236–255.

Zacks, J. M., S. Kumar, R. A. Abrams, and R. Mehta
2009. Using movement and intentions to understand human activity. *Cognition*, 112(2):201–216.

Zbikowski, L. M.
2002. *Conceptualizing music: cognitive structure, theory, and analysis*. Oxford ; New York: Oxford University Press.

Zima, E.
2014. English multimodal motion constructions. A construction grammar perspective. *Studies van de BKL – Travaux du CBL – Papers of the LSB*, 8:14–29.

Zima, E. and A. Bergs
2017. Multimodality and construction grammar. *Linguistics Vanguard*, 3(s1).

Zlatev, J.
2008. The co-evolution of intersubjectivity and bodily mimesis. In *Converging Evidence in Language and Communication Research*, J. Zlatev, T. P. Racine, C. Sinha, and E. Itkonen, eds., Pp. 215–244. Amsterdam: John Benjamins.

Zlatev, J.
2016. Turning back to experience in Cognitive Linguistics via phenomenology. *Cognitive Linguistics*, 27(4):559–572.

Zwaan, R. A.
2016. Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, 23(4):1028–1034.

List of Figures

List of Tables

List of Abbreviations

1	first person
3	third person
ACC	accusative
AdjP	adjective phrase
Adv	adverb
ASL	American Sign Language
AR	Asymmetric Redundancy
CA	Conversation analysis
CL	Cognitive Linguistics
COP	copula
CVPT	character viewpoint
CxG	Construction Grammar
DAT	dative
DEM	demonstrative
DET	determiner
DGS	Deutsche Gebärdensprache (*German Sign Language*)
DISTR	distributive
ERP	event-related potentials
F	feminine
fps	frames per second
FUT	future
GEN	genitive
GG	Generative Grammar
GP	Growth Point
HZJ	Hrvatski znakovni jezik (*Croatian Sign Language*)
INF	infinitive
IPFV	imperfective
IPH	Information Packaging Hypothesis
ISH	Integrated Systems Hypothesis
IS	information structure
LIS	Lingua dei Segni Italiana (*Italian Sign Language*)
M	masculine
NGT	Nederlandse Gebarentaal (*Dutch Sign Language*)
N	noun
NP	noun phrase
ÖGS	Österreichische Gebärdensprache (*Austrian Sign Language*)
OVPT	observer viewpoint
PFV	perfective
PJM	Polski język migowy (*Polish Sign Language*)
PL	plural
PREP	preposition
PRF	perfect
PRG	progressive
PST	past
PTCP	participle
RRG	Role and Reference Grammar

RT	reaction time
SE	standard error
SG	singular
SIMP	simple
SSL	Swedish Sign Language
TAM	tense–aspect–mood
TID	Türk İşaret Dili (*Turkish Sign Language*)
VP	verb phrase

Appendices

A. List of languages

language	ISO	genus	family
Afrikaans	afr	Germanic	Indo-European
American Sign Language	ase	Sign Languages	
Austrian Sign Language	asq	Sign Languages	
Central Alaskan Yupik	esu	Eskimo	Eskimo-Aleut
Croatian Sign Language	csq	Sign Languages	
Croatian	hrv	Slavic	Indo-European
Czech	ces	Slavic	Indo-European
Danish	dan	Germanic	Indo-European
Dutch	nld	Germanic	Indo-European
Dutch Sign Language	dse	Sign Languages	
Eastern Arrernte	aer	Central Pama-Nyungan	Pama-Nyungan
English	eng	Germanic	Indo-European
French	fra	Romance	Indo-European
German	deu	Germanic	Indo-European
German Sign Language	gsg	Sign Languages	
Guinea Kpelle	gkp	Western Mande	Mande
Guugu Yimidhirr	kky	Northern Pama-Nyungan	Pama-Nyungan
Hungarian	hun	Ugric	Uralic
Italian	ita	Romance	Indo-European
Italian Sign Language	ise	Sign Languages	
Japanese	jpn	Japanese	Japanese
Korean	kor	Korean	Korean
Mandarin Chinese	cmn	Chinese	Sino-Tibetan
Modern Greek	ell	Greek	Indo-European
Modern Hebrew	heb	Semitic	Afro-Asiatic
Nhengatu	yrl	Tupi-Guaraní	Tupian
Polish	pol	Slavic	Indo-European
Russian	rus	Slavic	Indo-European
Shona	sna	Bantoid	Niger-Congo
Siwu	akp	Kwa	Niger-Congo
Slovak	slk	Slavic	Indo-European
Spanish	spa	Romance	Indo-European
Standard Arabic	arb	Semitic	Afro-Asiatic
Suyá	suy	Ge	Macro-Ge
Swedish	swe	Germanic	Indo-European
Swedish Sign Language	swl	Sign Languages	
Turkish	tur	Turkic	Altaic
Turkish Sign Language	tsm	Sign Languages	
Tzeltal	tzh	Mayan	Mayan
Western Farsi	pes	Iranian	Indo-European
Yopno	yut	Finisterre-Huon	Trans-New Guinea

B. Informed consent

Filozofická fakulta Univerzity Karlovy

Souhlas s účastí ve výzkumu

Odpovědná osoba:

Identifikační číslo účastníka:

Prosím zaškrtněte

1. Potvrzuji, že jsem byl/a poučen/a o výzkumu a že jsem měl/a možnost položit doplňující otázky k účasti a k výzkumu samotnému.	
2. Jsem si vědom/a toho, že má účast ve výzkumu je dobrovolná a mám právo kdykoli odstoupit.	
3. Tímto dávám svolení členům výzkumného kolektivu ke zpracování svých anonymizovaných odpovědí v dotazníku. Jsem si vědom/a toho, že mé osobní údaje nebudou použity ve výstupech výzkumu.	
4. Souhlasím s pořízením audio- a video nahrávek zachycující mou osobu.	
5. Beru na vědomí, že pořízené nahrávky budou použity pouze v rámci výzkumu, za účelem prezentace jeho výsledků, příp. při výuce.	
6. Souhlasím se svou účasti ve výzkumu.	

_____	_____	_____
Jméno účastníka	Datum	Podpis
_____	_____	_____
Osoba odpovědná za provedení výzkumu	Datum	Podpis

Kopie:

1× účastník
1× archiv

C. Questionnaire

Dotazník

Identifikační číslo účastníka:[1]

Věk:

Pohlaví:

Ve kterém městě jste chodil/a do školy?

ZŠ:

SŠ:

Mateřský jazyk[2]:

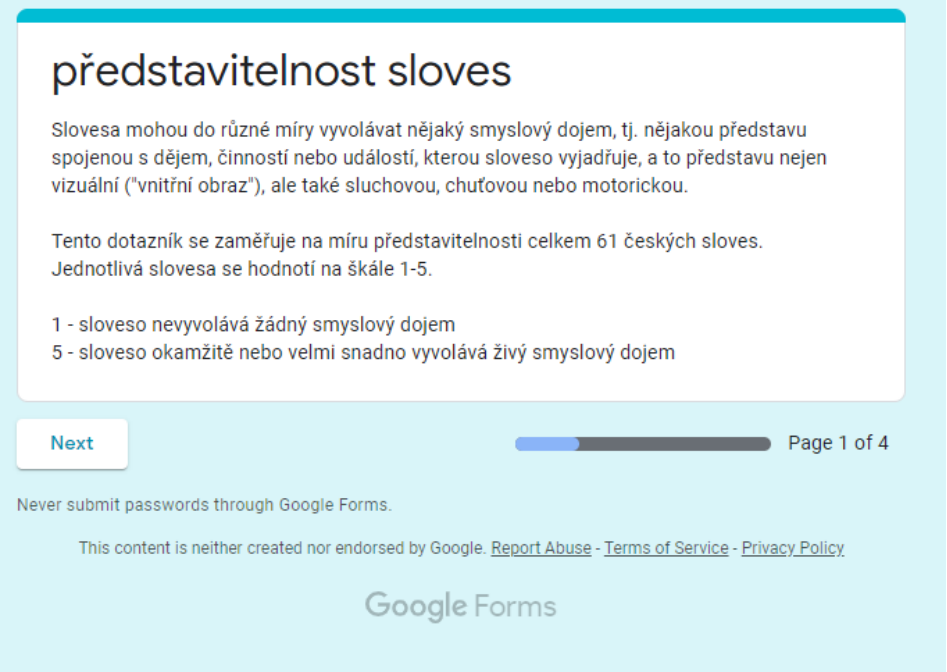
Další jazyky, které používáte/umíte/učíte se (včetně znakových):

1. (jak dlouho se jazyk učíte:)
2. (jak dlouho se jazyk učíte:)
3. (jak dlouho se jazyk učíte:)

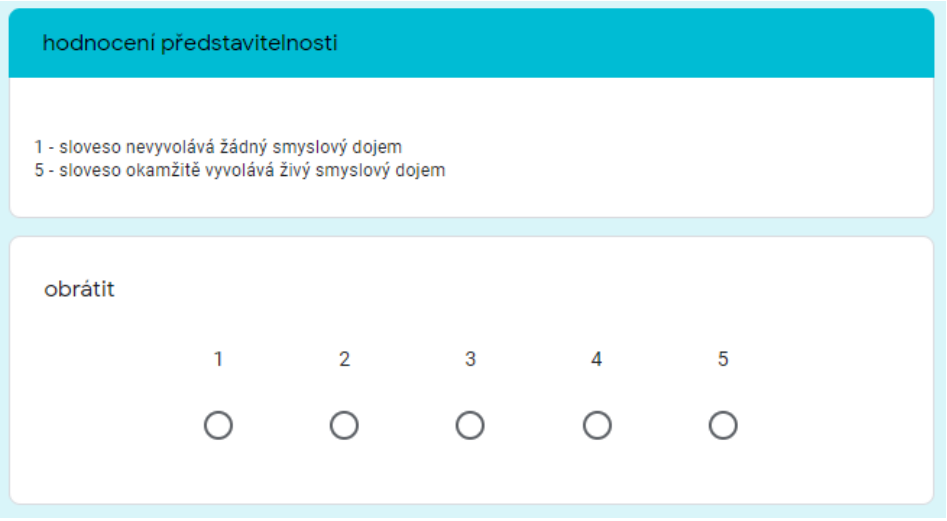
Kterou rukou píšete:

[1] Vyplní administrátor výzkumu
[2] Pokud jste bilingvní, můžete uvést více jazyků

D. Imageability rating form



Instructions



Answer sheet snippet

E. List of verbs and imageability ratings

lemma	translation	aspect	mean imageability	SD
brát	'take'	IPFV	3.18	1.25
číst	'read'	IPFV	3.92	0.79
dát	'give'	PFV	3.58	1.16
dodat	'add'	PFV	1.83	1.03
dostat	'get'	PFV	2.92	1.38
hodit	'throw'	PFV	4.50	0.90
koupit	'buy'	PFV	3.25	1.14
nabídnout	'offer'	PFV	2.50	1.45
najít	'find'	PFV	2.08	1.00
napadnout	'get an idea'	PFV	2.67	1.37
nechat	'let'	PFV	1.25	0.45
objevit	'discover'	PFV	1.92	0.90
obrátit	'turn over'	PFV	3.08	1.00
odmítnout	'refuse'	PFV	2.00	1.13
opustit	'leave'	PFV	2.17	1.11
otevřít	'open'	PFV	4.17	0.72
otočit	'turn around'	PFV	3.75	0.62
oznámit	'announce'	PFV	2.17	0.94
podat	'pass'	PFV	3.92	0.79
pochopit	'comprehend'	PFV	1.50	0.80
položit	'lay'	PFV	3.67	1.15
poslat	'send'	PFV	2.83	1.27
postavit	'build'	PFV	3.67	0.98
potkat	'encounter'	PFV	3.58	1.08
potvrdit	'confirm'	PFV	1.67	0.78
poznamenat	'note'	PFV	2.25	1.36
poznat	'get to know'	PFV	1.58	1.16
přijmout	'receive'	PFV	2.50	1.24
přinést	'fetch'	PFV	3.25	1.29
připravit	'prepare'	PFV	2.08	1.31
psát	'write'	IPFV	4.42	0.79
pustit	'drop'	PFV	3.17	1.40
rozhodnout	'decide'	PFV	1.25	0.45
říci	'say'	PFP	3.50	1.38
slyšet	'hear'	IPFV	3.92	1.16
udělat	'make'	PFV	2.25	1.42
ukázat	'show'	PFV	3.50	1.24
vidět	'see'	IPFV	3.75	1.36
vrátit	'return'	PFV	2.67	0.98
vstoupit	'enter'	PFV	4.17	0.58
vybrat	'choose'	PFV	2.17	0.83
vydat	'publish'	PFV	2.00	1.00
vyhrát	'win'	PFV	2.33	1.44
vyprávět	'narrate'	PFV	3.08	1.38
vyrazit	'knock out'	PFV	2.33	0.98

vysvětlit	'explain'	PFV	2.25	0.87
vytáhnout	'draw out'	PFV	3.75	1.22
vytvořit	'create'	PFV	2.25	1.36
začít	'begin'	PFV	1.50	0.90
zapomenout	'forget'	PFV	1.33	0.65
zastavit	'stop'	PFV	3.17	0.94
zavřít	'close'	PFV	4.17	0.94
zažít	'experience'	PFV	1.27	0.65
zdát se	'dream'	IPFV	2.08	1.31
získat	'gain'	PFV	1.67	0.98
zjistit	'find out'	PFV	1.17	0.39
změnit	'change'	PFV	1.33	0.65
znát	'know'	PFV	1.08	0.29
ztratit	'lose'	PFV	2.25	1.42
zvednout	'pick up'	PFV	4.00	1.04

F. Stimulus sentences

IPFV	PFV
Uruguayci bezpráví také zažívali.	Uruguayci bezpráví také zažili.
Něco si u stolu poznamenával.	Něco si u stolu poznamenal.
Babička na tetu zapomínala.	Babička na tetu zapomněla.
V katalogu si něco vybíral.	V katalogu si něco vybral.
Nezdravé návyky jsme opouštěli.	Nezdravé návyky jsme opustili.
Oni nám ty balíky budou dodávat.	Oni nám ty balíky dodají.
Radost ze života jsem ztrácel.	Radost ze života jsem ztratil.
My ty částky budeme rádi zjišťovat.	My ty částky rádi zjistíme.
Ty nabídky jsem odmítal.	Ty nabídky jsem odmítl.
Mistrovství v šachu začínalo.	Mistrovství v šachu začalo.
Ten zápas jsme vyhrávali.	Ten zápas jsme vyhráli.
Pro tebe jsem výjimky rád dělal.	Pro tebe jsem výjimky rád udělal.
Aritmetiku jsem dobře chápal.	Aritmetiku jsem dobře pochopil.
Svého psa doma nechávala.	Svého psa doma nechala.
Toho člověka jsem poznal.	Toho člověka jsem znal.
Povolení obtížně získávali.	Povolení obtížně získali.
Nová koncepce se bude vytvářet.	Nová koncepce se vytvoří.
Souvislosti nám poučeně vysvětloval.	Souvislosti nám poučeně vysvětlil.
Symptomy nemoci se měnily.	Symptomy nemoci se změnily.
O tom zákoně budou rozhodovat.	O tom zákoně rozhodnou.
Ty obědy se budou vydávat.	Ty obědy se vydají.
Na odjezd jsme se celkem připravovali.	Na odjezd jsme se celkem připravili.
Cestu k uznání si nějak nacházel.	Cestu k uznání si nějak našel.
Různá řešení nám nabízeli.	Různá řešení nám nabídli.

G. Gesture perception experiment

General information and instructions

Dobrý den,
na Filozofické fakultě UK provádíme online experiment, který je dostupný z této adresy: https://run.pavlovia.org/jakub.jehlicka/forcedchoice2/html – budeme rádi, když si najdete čas na jeho vyplnění. Nejde o nic složitého ani časově náročného. Experiment se zaměřuje na porozumění českým větám na základě vizuální informace. Po kliknutí na uvedený odkaz budete vyzváni k vyplnění stručného dotazníku. Své jméno nikde neuvádíte, Vaše účast je zcela anonymní.

Ovládání je pomocí klávesnice (mezerník a šipky), experiment lze spustit v prohlížeči Firefox v systémech Windows, macOS i Linux – nespouštějte jej prosím v chytrých telefonech. (Pokud jako výchozí prohlížeč používáte Chrome, experiment se spolehlivě podaří spustit jen v systému macOS, ve Windows proto prosím použijte prohlížeč Firefox).

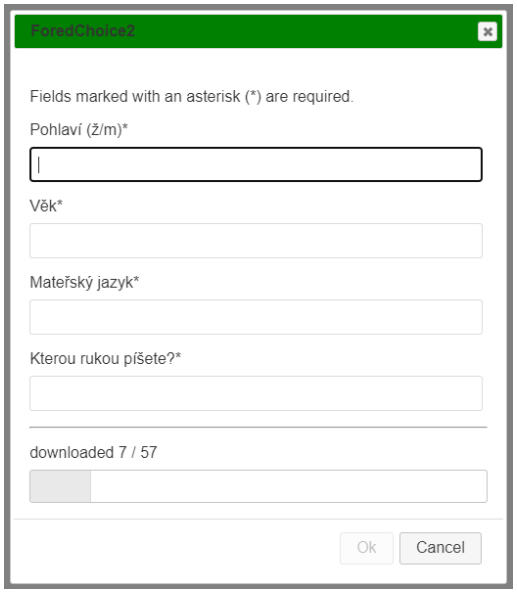
Během vyplňování údajů se experiment načte (během pár sekund v závislosti na rychlosti připojení), jakmile je vše připraveno, kliknete na „ok“ a následně se Vám zobrazí okno s instrukcemi, které si prosím pečlivě přečtěte. Samotný experiment zabere maximálně 10 minut.

Smysl experimentu nespočívá ve správných a špatných odpovědích – není to jazykový test, zajímá nás Vaše bezprostřední reakce. Proto nad odpověďmi příliš dlouho nepřemýšlejte a odpovídejte co nejrychleji. Experiment prosím vyplňte jen jednou.

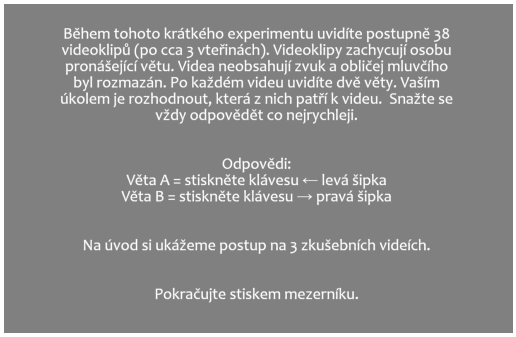
Děkuji Vám za spolupráci!

Jakub Jehlička
Ústav obecné lingvistiky
Filozofická fakulta Univerzity Karlovy
Mail: jakub.jehlicka@ff.cuni.cz

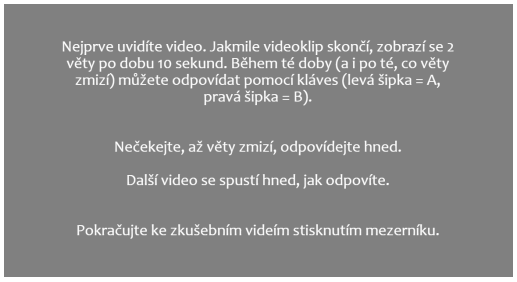
User interface snippets



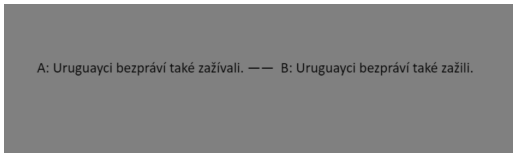
Participant information



Instruction screen 1



Instruction screen 2



Trial screen